

**A System-Wide Approach to Diabetic Nephropathy**

By

Luis Andrés de la Mora Palafox

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science in Biological Engineering

King Abdullah University of Science and Technology

Thuwal

Kingdom of Saudi Arabia

July 2011

The thesis **A System-Wide Approach to Diabetic Nephropathy** by Luis Andrés de la Mora Palafox is approved by:

Committee Chairperson: Dr. Timothy Ravasi

Committee Member: Dr. Jasmeen Merzaban

Committee Member: Dr. Christoph Gehring

© July 2011

*Luis Andrés de la Mora Palafox*

All Rights Reserved

This work is dedicated  
To my mother and my father

Ustedes me enseñaron  
A caminar entre los espacios  
Que nos separan del conocimiento

**ACKNOWLEDGEMENTS**

I want to thank Timothy Ravasi for giving me an opportunity to work with him. But especially I want to thank him for sharing his knowledge and philosophy of life with me.

Also, I want to thank Dr. Satish Rao and Dr. Kumar Sharma for been a guide and a support during my time in San Diego and across the whole process.

Finally, I want to thank the KAUST Graduate Skills Lab, in particular Dr. Ari Sherris for helping me finding the correct way to transfer my ideas into the paper.

The process has been long and difficult. I am sure that without the support of the people I mention here and many others that were next to me all the time I wouldn't have accomplished my goals.

**ABSTRACT****A System-Wide Approach to Diabetic Nephropathy**

Luis Andrés de la Mora Palafox

Diabetes mellitus is a complex human disease that affects more than 280 million people worldwide. One of the diabetic long-term complications is diabetic nephropathy that it is responsible for 50% of all end-stage renal disease. The complexity of diabetes and the lack of comprehensive systematic studies have halted the development of drugs and clinical therapies for the treatment of diabetes and its major complications. The present project, based on the *db/db* mice as animal model, investigates the repercussions of diabetes mellitus in the transcriptome as well as the mechanism of action of pirfenidone, an antifibrotic drug, in the treatment of diabetic nephropathy. The study was centered on the system-wide measurements transcriptional state of the mouse kidney. The expression profile of three experimental groups: control, diabetic, and diabetic treated with the drug, were analyzed using expression clustering, gene ontology enrichment analysis, protein-protein interaction network mapping, and gene expression behavior. The results show significant expression dysregulation of genes involved in RNA processing, fatty acid oxidation, and oxidative phosphorylation under the diabetic condition. The drug is able to regulate the expression levels of RNA processing genes but it does not show any effect in the expression profile of genes required in the oxidative phosphorylation and in the fatty acid metabolism. In conclusion diabetes mellitus induce the dysregulation of the splicing apparatus, the oxidative phosphorylation, and the fatty acid metabolic pathway at an expression level. The malfunction of these biological

pathways causes cellular stress by increasing the concentration of reactive oxygen species within the cell due to a high oxidative and respiratory activity of mitochondria in addition to the increased demand of the folding machinery as a consequence of a dysregulation of the splicing apparatus. Pirfenidone regulates the expression of RNA processing genes mainly by controlling the expression of peroxisome proliferator-activated receptor- $\gamma$  coactivator-1 $\alpha$ . The expression regulation overcomes the malfunction of the splicing apparatus and reduces the demand of the folding machinery. However the expression of genes annotated for fatty acid oxidation and oxidative phosphorylation do not change after drug treatment.

## TABLE OF CONTENTS

EXAMINATION COMMITTEE APPROVAL .....	2
COPYRIGHT .....	3
DEDICATION .....	4
ACKNOWLEDGEMENTS .....	5
ABSTRACT .....	6
TABLE OF CONTENTS .....	8
LIST OF TABLES .....	10
LIST OF FIGURES .....	11
LIST OF ABBREVIATIONS .....	14
Chapter I : Purpose.....	17
1.1 Motivation.....	17
1.2 Objectives .....	19
Chapter II : Introduction .....	20
2.1 Diabetes.....	20
2.1.1 Diabetic Nephropathy.....	25
2.1.2 The Diabetes Disease Models .....	26
2.1.3 The db/db Mouse Model.....	28
2.1.4 Diabetes and System Biology.....	28
2.2 System Biology .....	31
2.2.1 Bonding in Science.....	35
2.2.2 Computational Biology .....	36
2.2.3 Network Biology.....	38
Chapter III : Previous Work.....	41
Chapter IV : Results and Discussion .....	47
5.1 Preliminary Data Treatment.....	47
5.2 Statistical Analysis of Microarray .....	51
5.3 Human-Mouse Orthology .....	54
5.4 Gene Ontology Enrichment Analysis .....	56
5.5 Protein-Protein Interaction (PPI) Network .....	61
5.6 Peroxisome proliferator-activated receptor- $\gamma$ coactivator-1 $\alpha$ (PGC-1 $\alpha$ ).....	69
5.7 Fatty Acid Metabolic Pathway and Oxidative Phosphorylation.....	106



Chapter V : Materials and Methods .....	116
6.1 Experiment Design.....	116
6.2 RNA Isolation and Analysis .....	117
6.3 Microarray Data Preliminary Filtration .....	117
6.4 Statistical Analysis and Gene Expression Clustering .....	118
6.5 Orthology .....	119
6.6 Protein-Protein Interaction Network.....	119
6.7 Gene Ontology Enrichment Analysis .....	120
6.8 Expression Analysis.....	120
6.9 Fatty Acid Metabolism Pathway and Oxidative Phosphorylation.....	121
Chapter VI : Final Remarks .....	123
7.1 Conclusion .....	123
7.2 Limitations and Future Work.....	128
REFERENCES .....	130

**LIST OF TABLES**

Table 1 : Animal models of type 1 diabetes .....	27
Table 2 : Animal models of type 2 diabetes .....	27
Table 3 : Enrichment analysis of the expression clusters .....	57
Table 4 : Major network hubs .....	61
Table 5 : Genes part of the proteasome complex significant expressed in the diabetic and treated group.....	66
Table 6 : PGC-1 $\alpha$ first neighbors .....	69
Table 7 : Enrichment analysis of PGC-1 $\alpha$ and its neighbors .....	74
Table 8 : Functional clusters (FC) for PGC-1 $\alpha$ neighbors .....	80
Table 9 : Enrichment analysis of SFRS4, SFRS5, SFRS5, and CPSF2 .....	85
Table 10 : Enrichment analysis of PPAR $\alpha$ , PPAR $\gamma$ and their first neighbors .....	90
Table 11 : Enrichment analysis of NRF-1, NRF-2 and their first neighbors .....	100
Table 12 : Enrichment analysis of TFAM and its first neighbors.....	104
Table 13 : Significant expressed genes annotated in lipid metabolism pathway .....	107
Table 14 : Mitochondrially encoded genes annotated in the oxidative phosphorylation.....	108

## LIST OF FIGURES

Figure 1 : Worldwide diabetes prevalence for 2010 .....	22
Figure 2 : Eastern Mediterranean and Middle East Region diabetes prevalence for 2010.....	23
Figure 3 : Chemical structure of PFD .....	42
Figure 4 : Glucose-induced signaling pathway in mesangial cells .....	44
Figure 5 : Intensity vs. RNA concentration correlation .....	48
Figure 6 : $f_{2D}$ as a function of inter-probe spacing ( $R_b$ ) for various target size .....	49
Figure 7 : Hybridization of fluoresceine-labeled oligonucleotide on a microchip .....	50
Figure 8 : Venn diagram of the expression clusters.....	51
Figure 9 : <i>Mus musculus</i> gene expression clusters .....	52
Figure 10 : Relevant expression clusters .....	53
Figure 11 : <i>Homo sapiens</i> orthologous gene expression clusters .....	55
Figure 12 : Expression profile of genes significant expressed in the control and treated expression cluster annotated in RNA processing.....	60
Figure 13 : Control PPI subnetwork .....	63
Figure 14 : Diabetic PPI subnetwork .....	64
Figure 15 : Treated PPI subnetwork .....	64
Figure 16 : Control and Diabetic PPI subnetwork .....	65
Figure 17 : Control and Treated PPI subnetwork .....	65
Figure 18 : PPI subnetwork enriched in protein metabolic complex part of the diabetic and treated PPI subnetwork .....	66
Figure 19 : Diabetic and Treated PPI subnetwork .....	68
Figure 20 : Expression profile of PGC-1 $\alpha$ .....	71
Figure 21 : PPI subnetwork centered on PGC-1 $\alpha$ (PPARGC1A) .....	76
Figure 22 : Expression profile of CCNT1 and EP300 and correlation with PGC-1 $\alpha$ ..	77
Figure 23 : Expression profile of ESR1 and ESRRG and correlation with PGC-1 $\alpha$ ..	77
Figure 24 : Expression profile of NRF1 and PPARGC1A and correlation with PGC-1 $\alpha$ ..	78
Figure 25 : Expression profile of MED1 and SFRS4 and correlation with PGC-1 $\alpha$ ..	78
Figure 26 : Expression profile of SFRS5 and SFRS6 and correlation with PGC-1 $\alpha$ ..	79
Figure 27 : Expression profile of USF2 and MED24 and correlation with PGC-1 $\alpha$ ..	79

Figure 28 : Expression profile of NR1H4 and NCOA6 and correlation with PGC-1 $\alpha$ .....	80
Figure 29 : Architecture of PGC-1 $\alpha$ protein.....	82
Figure 30 : Expression profile of SFRS5 and SFRS6.....	83
Figure 31 : Expression profile of SFRS4.....	83
Figure 32 : Network hubs within the PGC-1 $\alpha$ subnetwork .....	84
Figure 33 : SFRS4, SFRS5, SFRS6, and CPSF2 common neighbors .....	85
Figure 34 : PPI subnetwork of RNA processing genes.....	86
Figure 35 : Expression profile of genes involved in RNA processing .....	87
Figure 36 : PPI subnetwork centered in dysregulated genes involved in RNA processing .....	87
Figure 37 : Expression profile of PPAR $\alpha$ .....	89
Figure 38 : PPI subnetwork centered in PPAR $\alpha$ and PPAR $\gamma$ .....	89
Figure 39 : Adipocytokines Signaling Pathway.....	92
Figure 40 : Expression profile of RXR relative to PPAR $\alpha$ .....	93
Figure 41 : Expression profile of ESR1 .....	94
Figure 42 : PPI subnetwork centered in the ESRs and ESRRs.....	95
Figure 43 : Expression profile of ESRR $\alpha$ , ESRR $\beta$ and ESRR $\gamma$ .....	96
Figure 44 : Expression correlation between ESR1 and ESRR $\alpha$ .....	97
Figure 45 : Diagram of the expression of mitochondrial genes.....	98
Figure 46 : Expression profile of NRF-1 .....	99
Figure 47 : PPI subnetwork centered in NRF-1 and NRF-2 .....	100
Figure 48 : Schematic representation of the human mitochondrial genome .....	101
Figure 49 : Expression profile of TFAM and expression correlation with PGC-1 $\alpha$ ..	102
Figure 50 : PPI subnetwork centered in TFAM.....	103
Figure 51 : Expression profile of TF1BM and TF2BM.....	105
Figure 52 : Expression profile of POLRMT .....	106
Figure 53 : PPI subnetwork of genes annotated in the lipid metabolism pathway and their first neighbors .....	109
Figure 54 : Dysregulated genes annotated in lipid metabolism pathway .....	110
Figure 55 : Dysregulated genes annotated in lipid metabolism pathway .....	111
Figure 56 : PPI subnetworks built with the genes annotated in oxidative phosphorylation.....	112

Figure 57 : Dysregulated genes annotated in oxidative phosphorylation pathway ...113

Figure 58 : Dysregulated genes annotated in oxidative phosphorylation pathway ...114

**LIST OF ABBREVIATIONS**

<b>ADA</b>	American Diabetes Association
<b>ADIPO</b>	adiponectine
<b>ADIPOR</b>	adiponectin receptors
<b>AIDS</b>	acquired immune deficiency syndrome
<b>BCS1L</b>	BCS1-like ( <i>S. cerevisiae</i> )
<b>BISTI</b>	Biomedical Information Science and Technology Initiative
<b>CCNT1</b>	cyclin T1
<b>cDNA</b>	complementary DNA
<b>ChIP</b>	chromatin immunoprecipitation
<b>coIP</b>	co-immunoprecipitation
<b>CPSF2</b>	cleavage and polyadenylation specific factor 2, 100kDa
<b>CREBBP</b>	CREB binding protein
<b>DM</b>	diabetes mellitus
<b>DN</b>	diabetic nephropathy
<b>DNA</b>	deoxyribonucleic acid
<b>ECM</b>	extracellular matrix
<b>EP300</b>	E1A binding protein p300
<b>ER</b>	endoplasmic reticulum
<b>ESR</b>	estrogen receptor
<b>ESR1/ER<math>\alpha</math></b>	estrogen receptor 1
<b>ESR2</b>	estrogen receptor 2
<b>ESRD</b>	end-stage renal disease
<b>ESRR</b>	estrogen-related receptor
<b>ESRR<math>\alpha</math></b>	estrogen-related receptor alpha
<b>ESRR<math>\beta</math></b>	estrogen-related receptor beta
<b>ESRR<math>\gamma</math>/ESRR<math>\gamma</math></b>	estrogen-related receptor gamma
<b>F.E.</b>	fold enrichment
<b>FDR</b>	false discovery rate
<b>FFA</b>	free fatty acid
<b>GDM</b>	Gestational Diabetic Mellitus
<b>GFR</b>	glomerular filtration rate
<b>HAT</b>	histone acetyl transferase
<b>HNF-1<math>\alpha</math></b>	hepatocyte nuclear factor-1 $\alpha$
<b>HNF4A</b>	hepatocyte nuclear factor 4, alpha
<b>IDDM</b>	Insulin-Dependent Diabetes Mellitus
<b>IDF</b>	International Diabetes Federation
<b>IRS</b>	insuline receptor substrate
<b>IRS-1</b>	insulin receptor substrate 1
<b><i>k</i></b>	degree
<b>LC-MS/MS</b>	liquid chromatography with subsequent tandem mass

	spectrometry
<b>MAPK</b>	mitogen-activated protein kinases
<b>MED24</b>	mediator complex subunit 24
<b>MGI</b>	Mouse Genome Informatics
<b>MHRE</b>	multi hormone response element
<b>MC</b>	mesangial cells
<b>MMC</b>	murine mesangial cells
<b>MODY</b>	maturity-onset diabetes of the young
<b>mRNA</b>	messenger RNA
<b>mtDNA</b>	mitochondrial DNA
<b>MTF1</b>	metal-regulatory transcription factor 1
<b>NDDG</b>	National Diabetes Data Group
<b>NDM</b>	neonatal diabetes mellitus
<b>NIDDK</b>	National Institute of Diabetes and Digestive and Kidney Disease
<b>NIDDM</b>	Non-Insulin-Dependent Diabetes Mellitus
<b>NIH</b>	U.S. National Institute of Health
<b>NMR</b>	nuclear magnetic resonance
<b>NR1H4</b>	nuclear receptor subfamily 1, group H, member 4
<b>NRF</b>	nuclear respiratory factor
<b>NRF1</b>	nuclear respiratory factor 1
<b>NRF2</b>	nuclear respiratory factor 2
<b>OXPHOS</b>	oxidative phosphorylation
<b>PCA</b>	protein-fragment complementation assay
<b>PFD</b>	pirfenidone
<b>POLR2A</b>	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa
<b>POLRMT</b>	polymerase RNA mitochondrial
<b>PPAR</b>	peroxisome proliferator-activated receptors
<b>PPARA/PPAR<math>\alpha</math></b>	peroxisome proliferator-activated receptor alpha
<b>PPAR<math>\gamma</math></b>	peroxisome proliferator-activated receptor gamma
<b>PPARGC1A/PGC-1<math>\alpha</math></b>	Peroxisome proliferator-activated receptor- $\gamma$ coactivator-1 $\alpha$
<b>PPI</b>	protein-protein interaction
<b>PRC</b>	PGC-1-related coactivator
<b>R</b>	Pearson's correlation coefficient
<b>RMM</b>	RNA-binding motif
<b>RNA</b>	ribonuclein acid
<b>ROS</b>	reactive oxygen species
<b>rRNA</b>	ribosomal RNA
<b>RT-PCR</b>	reverse transcription polymerase chain reaction
<b>RXR</b>	retinoid X receptor
<b>SFRS4</b>	serine/arginine-rich splicing factor 4
<b>SFRS5</b>	serine/arginine-rich splicing factor 5

<b>SFRS6</b>	serine/arginine-rich splicing factor 6
<b>SMAD2</b>	SMAD family member 2
<b>SNP</b>	single nucleotide polymorphisms
<b>SNRPG</b>	small nuclear ribonucleoprotein polypeptide G
<b>SP1</b>	specificity protein 1
<b>TAIR</b>	<i>Arabidopsis</i> Information Resource
<b>TetR</b>	tetracycline repressor
<b>TFAM</b>	transcription factor A, mitochondrial
<b>TFB1M</b>	transcription factor B1, mitochondrial
<b>TFB2M</b>	transcription factor B2, mitochondrial
<b>TGF-<math>\beta</math>1</b>	transforming growth factor $\beta$ 1
<b>TNF-<math>\alpha</math></b>	transforming growth factor $\alpha$
<b>TR</b>	thyroid hormone receptors
<b>tRNA</b>	transfer RNA
<b>UCP</b>	mitochondrial uncoupling proteins
<b>UPR</b>	unfolded protein response
<b>USF2</b>	upstream transcription factor 2
<b>WAT</b>	white adipose tissue
<b>WHO</b>	World Health Organization



## **Chapter I : Purpose**

### **1.1 Motivation**

In 2010 around 7% of the total world population was diagnosed with diabetes. By 2030 the worldwide diabetic population will increase by 50% affecting more than 430 million people. Diabetes is a social, economic, and health problem around the world that requires special attention. However the treatment of diabetes and its major complications is difficult due to the great amount of genetic, environmental and biological events that play a part in the development of the disease.

Complex human diseases, such as, arthritis (Attur et al. 2002) acquired immune deficiency syndrome (AIDS) (Doolittle & Gomez 2010), cancer (Ferreira et al. 2010), Alzheimer (J. A. Miller et al. 2010), and diabetes (Mullen & Ohlendieck 2010) have been explored using systematic approaches. The onset and development of these diseases involves the malfunction of multiple biological mechanisms, which make it difficult to completely characterize the disease. The complexity of diabetes derives from the multifactorial events that trigger the disease and the dysregulation of several biological pathways and functions at different levels during development.

A system-wide approach is required in order to have an entire panoramic of the impact of diabetes in the function of a complete system. The lack of expression-based analyses in diabetic models has prevented the complete description and characterization of the pathogenesis and development of diabetes, as well as its major complications.

Since the introduction of systems biology into science, it has been clear that the overall behavior of an organism, a disease, or a signaling pathway needs to be explained taking into consideration all the elements that may participate in a specific event, the interactions between these elements, and the dynamics of adaptation and response. It is evident that every biological event is the result of many physicochemical phenomena interacting together and the final outcome is the average effect of all these incidents and its interactions. System biology has allowed researchers to extend their investigation into new directions that years ago seemed unexplorable. Thanks to cutting-edge technology on the fields of genomics, bioinformatics, and genetics it is now possible to identify, measure, and quantify the multiple events that function together in any biological workflow.

The goal of this proposal is to perform a complete expression-based analysis of the *db/db* mice. This study specifically sets forth to properly analyze the microarray data in combination with protein-protein interaction networks. The integration of these approaches will identify the main biological pathways that have been disrupted and will allow the understanding of the overall behavior of the disease within a complete system. It is expected that the system-wide approach will provide a new perception of diabetes that will become essential in developing new and novel therapies for diabetes.

## 1.2 Objectives

The present project studies the expression profile in kidney tissue of the *db/db* mouse with and without the treatment of pirfenidone (PFD). The study is performed using a complete expression-based network analysis with the objective to identify biomarkers of the disease and infer the mechanism of action of PFD.

Hence the study is centered on the search of an answer, at an expression level, to the following questions.

1) Which biological pathways are being affected by the presence of type 2 diabetes mellitus?

2) What is the mechanism of action of PFD in the treatment of diabetic nephropathy?

## **Chapter II : Introduction**

### **2.1 Diabetes**

Diabetes mellitus (DM) is a clinically and genetically heterogeneous group of disorders that are characterized by high levels of glucose in the blood due to insulin deficiency, insulin resistance, or both (American Diabetes Association 2011). There are four major types of diabetes according to the classification proposed by the National Diabetes Data Group (NDDG) and the World Health Organization (WHO) (The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus 2002):

1. Type 1 Diabetes or Insulin-Dependent Diabetes Mellitus (IDDM): results from a cellular-mediated autoimmune destruction of pancreatic  $\beta$ -cells (Grieco et al. 2011). The patients lack the capacity to produce insulin and required to be treated with insulin. About 5-10% of all diabetic patients suffer from type 1 diabetes mellitus (American Diabetes Association 2011).
2. Type 2 Diabetes or Non-Insulin-Dependent Diabetes Mellitus (NIDDM): many causes are associated with the onset of this type of diabetes. As a common denominator, most patients that present type 2 diabetes mellitus are obese; those patients that are not obese present an increased percentage of body fat in the abdominal region (Anderson et al. 2003). Patients show insulin resistance and usually have relative insulin deficiency. Type 2 diabetic

patients represent 90-95% of the total diabetic population (American Diabetes Association 2011).

3. Gestational Diabetic Mellitus (GDM): is defined as any degree of glucose intolerance recognized during pregnancy. Most cases resolve after delivery (Jovanovic 2009).
4. Other types of diabetes: some cases of diabetes are associated with specific conditions, such as, pancreatic disease, hormonal disease, drug or chemical exposure, insulin receptor abnormalities, and genetic syndromes.

Even though the different types of DM are triggered by distinct conditions their pathological courses after onset are similar. Long-term complications of DM include retinopathy with potential loss of vision, renal failure cause by nephropathy, atherosclerotic cardiovascular disease, hypertension, and heart failure (American Diabetes Association 2011).

Social factors, such as, population growth, aging, and urbanization, together with an increasing prevalence of obesity and physical inactivity are rapidly increasing the number of people with diabetes (Wild et al 2004). The dramatic rise in diabetic patients worldwide has turned this disease into an epidemic. In 2005 it was estimated that diabetes together with other chronic, noncommunicable diseases (e.g., cancer and cardiovascular disease) caused 35 million deaths. Moreover, by 2015 this death rate is expected to increase 17% (International Diabetes Federation 2009). A study considering statistics from 91 countries (the total population of 80 of these countries

represent 95% of the world adult population in 2010) estimates a total of 285 million diabetic patients worldwide in 2010 and an average increase by 54% over the next 20 years. In terms of development indicators this represents an average increase of 69% for developing countries and 20% for developed countries. The studies suggest a 2.2% annual growth rate of the worldwide diabetic population; which is nearly twice the annual growth of the total world adult population. This projection leads to 154 million new cases and a total worldwide population with diabetes of 439 million for 2030 (J. E. Shaw et al. 2010).

The 2010 world prevalence of diabetes (Figure 1) within the adult population (aged 20-79 years) was estimated of 6.4%. With the highest prevalence occurring in the North American Region (10.2%) and in the Eastern Mediterranean and Middle East Region (9.3%). While a 42.4% increase in the total number of adults with diabetes is expected for the North American Region for 2030, whereas the Eastern Mediterranean and Middle East Region will experience a 93.9% increase and a total of 51.7 million adults with diabetes for the same year.

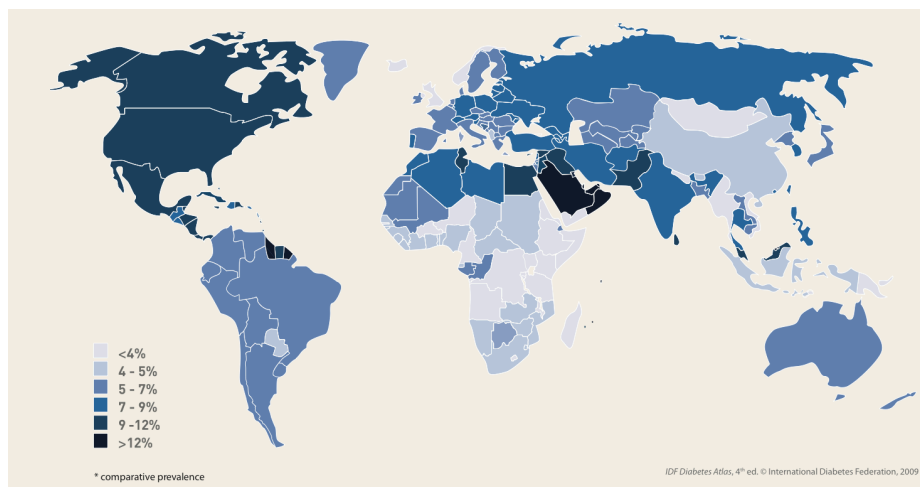


Figure 1  
Worldwide diabetes prevalence for 2010  
(International Diabetes Federation 2009)

Five of the top 10 countries with the highest national prevalence in 2010 belong to the Eastern Mediterranean and Middle East Region (UAE, Saudi Arabia, Bahrain, Kuwait, and Oman) (Figure 2). Saudi Arabia is in third place with a prevalence of 16.8% and an estimated of 2 million adults with diabetes with an annual growth of 100 thousand new cases during the next 20 years (J. E. Shaw et al. 2010).

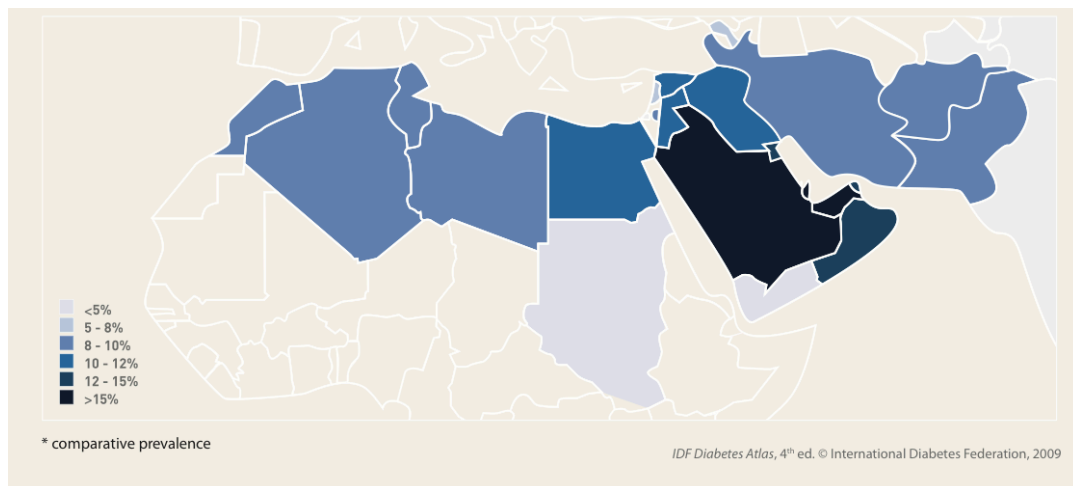


Figure 2  
Eastern Mediterranean and Middle East Region diabetes prevalence for 2010  
(International Diabetes Federation 2009)

Recent studies in the Saudi population suggest a prevalence of diabetes of 34.1% in males and 27.6% in females and a continuous rise in the prevalence rates over the next 20 years (Alqurashi et al. 2011). The International Diabetes Federation (IDF) reported in the 4<sup>th</sup> edition of the Diabetes Atlas, an estimate of 7,798 deaths in males and 6,307 deaths in females in Saudi Arabia attributable to diabetes in 2010. This represents about 23% of all mortality cause in females and 10% in males (International Diabetes Federation 2009). The Kingdom of Saudi Arabia is rapidly moving towards urbanization and its population is drastically changing their lifestyle. Different studies suggested that diabetes has become an epidemic in the country,

reaching high rates, above the world median, particularly in urban areas (Al-Nozha et al. 2004).

Diabetes has a great impact on the world economy and in the national healthcare system. Worldwide healthcare expenditures on diabetes estimates represent 11.6% of the total healthcare budget in 2010. The total cost to prevent and treat diabetes complications was evaluated in 376 billion US dollars for the same year, an average of 703 US dollars per patient (International Diabetes Federation 2009). According to the American Diabetes Association (ADA) the USA spent 174 billion US dollars in 2007 for mitigating the impact of diabetes, divided into 116 billion US dollars for direct medical care and 58 billion for indirect costs including lost productivity (American Diabetes Association 2008). It was estimated that 1 of each 5 US dollars dedicated to health care in the USA were allocated to people with diabetes (Ariza et al. 2010). The global cost by 2030 is predicted to exceed 490 billion US dollars (International Diabetes Federation 2009).

The Kingdom of Saudi Arabia spent 21% of the 2010 total healthcare expenditure in the treatment of diabetes. This represents about 2 billion US dollars and an average of 868 US dollars per patient. It is expected that by 2030 the total cost of diabetes in Saudi Arabia will be around 5 billion US dollars (International Diabetes Federation 2009).

It is clear that diabetes is a worldwide problem affecting many aspects of the social structure. Many organizations around the world are dedicating enormous effort in understanding the causes, behavior, and development of this disease in order to



prevent future cases and reduce the impact in current patients. The forecast for the coming years encourages researchers to work harder. System biology nowadays is providing new clues for combating this epidemic.

### ***2.1.1 Diabetic Nephropathy***

Diabetic nephropathy is an end-stage renal disease (ESRD) characterized by an increase albumin excretion and mesangial expansion that progress to sclerosis of glomerular capillaries (Shumway & Gambert 2002). The pathogenesis of diabetic nephropathy results from multiple interaction between metabolic pathways affected by environmental and genetic factors (Sanchez & Sharma 2009).

Diabetes mellitus is the leading cause of ESRD; about one third of the diabetic population develops a chronic kidney disease (Ha et al. 2008). The National Institute of Diabetes and Digestive and Kidney Disease (NIDDK)<sup>1</sup> estimates that 11.5% of the U.S. adult population has physiological evidence of chronic kidney disease. In 2007 111,000 new cases of ESRD were diagnosed and about 45% of these new cases were attributable to diabetes.

The lack of a comprehensive study of the etiology and mechanisms responsible for diabetic nephropathy has impeded the development of clinical strategies that prevent and cure this chronic kidney disease. Until today, there is not a clinical therapy that cures diabetic nephropathy. Nevertheless the clinical course of this disease can be control by changing the life style of the patients and with the use

---

<sup>1</sup> <http://www2.niddk.nih.gov/>

of new pharmacological therapies. The present project based on expression changes tries to understand the mechanism of action of pirfenidone (PFD) in the treatment of diabetic nephropathy.

### ***2.1.2 The Diabetes Disease Models***

The understanding of the development of DM and its complications has been halted due to the lack of reliable animal models that mimic human disease (Brosius et al. 2009). An adequate animal model that presents pathological and clinical characteristics similar to that of human would help in revealing the mechanisms behind DM and in testing new therapeutic drugs (Tamrakar et al. 2009).

Animal models for both, type 1 and type 2 diabetes, have been developed for more than 50 years. The efforts have focused on the mouse model due to the great amount of genomic information available and the standardized techniques for genetic manipulation that may be performed in mice (Hsueh et al. 2007). These models share great similarity to human disease and have been an important instrument for the study of diabetes.

The models for type 1 DM (Table 1) are characterized by the immunological destruction of  $\beta$ -cells and the development of absolute insulinopaenia (Rees & Alcolado 2005).

Table 1  
Animal models of type 1 diabetes

---

NOD (non-obese diabetic) mouse
BB (bio breeding) rat
LETL (Long Evans Tokushima lean) rat
New Zealand white rabbit
Keeshond dog
Chinese hamster
Celebes black ape ( <i>Macacca nigra</i> )

---

The animal models of type 2 diabetes (Table 2) tend to be as genetically complex and heterogeneous as the human condition, in some animals, insulin resistance predominates while in others  $\beta$ -cell failure is the cause of the disease (Srinivasan & Ramarao 2007).

Table 2  
Animal models of type 2 diabetes

---

Ob/Ob mouse - monogenic model of obesity (leptin deficient)
db/db mouse - monogenic model of obesity (leptin resistant)
Zucker (fa/fa) rat - monogenic model of obesity (leptin resistant)
Goto Kakizaki rat
KK mouse
NSY mouse
OLETF rat
Israeli sand rat
Fat-fed streptozotocin-treated rat
CBA/Ca mouse
Diabetic Torri rat
New Zealand obese mouse

---

### **2.1.3 The *db/db* Mouse Model**

The *db/db* mouse is a well-established animal model for DM and insulin resistance and has been used extensively for the research of type 2 diabetes. Fatty acid accumulation (M. Li et al. 2010), mitochondrial proteomics analysis in the diabetic state (Essop et al. 2010), and the effect of novel drugs in the treatment of diabetes (Yoshida et al. 2010) are some of the investigations that are using the *db/db* mouse as a research model.

Diabetes is triggered on the *db/db* mouse due to an autosomal recessive mutation in the *db* gene, which encodes for the leptin receptors (Srinivasan & Ramarao 2007). During the first month of life these mice become obese and insulin resistant. As a result, they resemble non-insulin-dependent diabetes mellitus in human. Moreover, between the 3-4 months of age they develop hyperinsulinemia and hyperglycemia (Tamrakar et al. 2009), and do not survive longer than 8-10 months.

### **2.1.4 Diabetes and System Biology**

The genetic analysis of DM has led to the discovery of many risk loci. This genetic knowledge has proved that the onset and development of diabetes is caused by multiple loci with specific effects that have not been well established. Most of the risk alleles are found in the patients, however in very few cases the patients are homozygous at a given locus. It is estimated that each individual risk locus increases the risk of diabetes by 5 to 40% (Meigs 2009). Common variants with large effects is not the genetic pattern in diabetes. Moreover diabetes results from the failure, at

multiple levels, of many regulatory and signaling pathways, such as, energy homeostasis, fatty acid oxidation, and hormone signaling. All these elements weave the complexity of diabetes making difficult to understand it based on a traditional biology focus. A traditional approach can lead to a limited understanding of the disease due to its reductionist angle of study. However, system biology by studying all the components and its interactions gives a wider perspective of how the disease operates, how can it be predicted and prevented, and how its health problems can be mitigated.

The first genetic studies of diabetes were based on single gene mutations. These studies helped in the definition of the etiology of diabetes. Early research works identified point mutations in the insulin receptor gene (T. Kadowaki et al. 1988) providing important evidence to the etiology of NIDDM. Other mutations in genes that encode for hepatocyte nuclear factor-1 $\alpha$  (HNF-1 $\alpha$ ) (Yamagata et al. 1996), glucokinase (Froguel et al. 1992), and peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ) (Barroso et al. 1999) were also identified using a single gene approach. These gene discoveries provided important information regarding the general aspects and mechanisms of diabetes mellitus. At the same time the limitations of the single gene analysis in the identification of compensatory pathways was highlighted. Simultaneously the single gene approach proved to be efficient for the diagnosis, the explanation of the clinical features, and the prediction of the clinical course of monogenic diabetes (McCarthy & Hattersley 2008), such as, neonatal diabetes mellitus (NDM) and maturity-onset diabetes of the young (MODY).

Monogenic forms of diabetes represent about 1 to 5 percent off all cases of diabetes (National Diabetes Information Clearinghouse 2007). The vast majority of the diabetes patients suffer from the polygenic forms of diabetes: type 1 and type 2 diabetes. The polygenic forms, especially type 2 diabetes, are influenced by genetic variations at multiple loci, environmental factors, and the regulation of several cellular mechanisms (Stumvoll et al. 2005). The challenge lies in achieving a deep mechanistic knowledge of the polygenic forms of the disease in order to provide clinical benefits to the hundreds of millions of people affected and at risk of type 1 and type 2 DM (McCarthy & Hattersley 2008).

The current understanding of the pathophysiology and the genetic background of diabetes is insufficient leading to inadequate therapeutic options and a poor description of the mechanisms responsible for the onset and the progression of the disease. The limited biological comprehension of diabetes is a consequence of its complexity, its polygenic characteristic, and its multifactorial traits. The genetic and biological analysis of diabetes requires a system biology approach in order to attain a deeper insight of the diseases. With the help of system biology many genes can be study at the same time enabling the identification of interaction networks that may be useful in the overall description of DM.

A system biology-driven approach is implemented in the present work for the analysis of diabetes mellitus. Molecular interaction and gene expression data are integrated yielding a complete investigation of the cellular mechanisms that may have an important role in the progression of the disease. This work uses the advantages of system biology and provides a comprehensive study of the genetic structure of DM.

## 2.2 System Biology

In the early years of the twentieth-century biology faced methodological problems that went beyond scientific and technological restrictions. It confronted the essential concerns stipulated by Ludwing von Bertalanffy in 1950 in his paper *An Outline of General System Theory* where he stated,

“[...] we can isolate processes occurring in the living organism and describe them in the terms and laws of physico-chemistry. This is done, with enormous success, in modern biophysics and biochemistry. But when it comes to the properly ‘vital’ features, it is found that there are essentially problems of organization, orderliness, and regulation, resulting from the interactions of an enormous number of highly complicated physico-chemistry events.” (Bertalanffy 1950)

Von Bertalanffy reframed science in order to address complex, adaptive, and nonlinear systems. He centered his approach in terms of “problems of organization”. It was evident that the complexity of biology could not be handled by the isolation of the components. There was a marked obligation of studying the parts and its relations, the ‘wholeness’ and its dynamic behavior. As a result of von Bertalanffy’s critique, the idea of analyzing biology from a systematic point of view was introduced.

Von Bertalanffy’s approach was hindered by the technological constraints of his time that prevented biology from entering the organismic methodology of today.

It was until the end of the nineteen nineties and the beginning of the third millennium when high-throughput data, technologies for characterization of cellular response, and complex mathematical algorithms were available that biologists and researchers gradually advance towards a system-level methodology (Kitano 2002b) and system biology was introduced into the concepts of modern biology.

After many years of scientific progress biology was allowed to enter a new era, shifting from the study of elements to the study of networks, from matter to states, and from structures to dynamics (Kitano 2002b). Finally the idea of a new scientific doctrine of ‘wholeness’ proposed by von Bertalanffy in his General System Theory developed so that it could provide a novel perspective on the intricacy of life through biological study, investigation, and analysis.

The transformation of the study of biology from a mechanistic/individualistic focus to a systematic/wholistic one has revolutionized the design of recent research studies. This new approach has lead to many important scientific developments and to the description of complex systems, such as, the Human Genome Project (Collins et al. 2003). Nevertheless, the complete characterization of complex systems will always remain at a philosophical level and in principal it is only possible to aspire to its understanding. The systematic approach introduced by von Bertalanffy has as main principle; the study of the dynamic behavior of life, its component elements, and the relations of forces between them with the objective of generating hypothetico-deductive models.



System biology investigates and describes the behavior and relationships of all the components of a particular biological system while it is functioning (Ideker et al. 2001). It tries to understand how the components assemble and how they respond to specific perturbations. It is a dynamic study of the system rather than just a structural analysis of the elements, the system is an abstract concept, is more than the simple sum of its parts. The complete evaluation of biology at a system-level consists of four key properties (Kitano 2002a):

1. System structures: a network of gene regulations and biochemical pathways, a organizational analysis that covers the elements of the network, interactions between elements, and the basic organization of the network.
2. System dynamics: the response of the network over time and under different environmental, chemical and physical conditions.
3. System control: mechanisms that regulate the state of the network and minimize malfunctions.
4. System design: the development and formulation of specific biological systems with desired properties and functions.

System biology deals with the first three areas while synthetic biology is in charged of the modeling and design of new systems. The organizational approach of system biology attempts to understand the working of biological networks and

synthetic biology uses this knowledge to construct new genetic and biological systems (H.-Y. Chuang et al. 2010).

Biological systems are dynamic entities with a continuous flux of information that determinates the equilibrium and correct function of the system and its relation with the environment. This information is stored in two main packages: genes and regulatory interactions. The information has a hierarchical motion that moves from DNA to mRNA, followed by protein and protein interactions (e.g., informational pathways and networks) (Ideker et al. 2001). The responsibility of system biology is to collect data from each of these levels and generate predictive models that can simulate the flow of the information and the functioning of the system. Even though the transition of biology into a new systematic conception is a consequence of the ability to obtained genome, transcriptome and proteome large-scale data. System biology is not only about collection and storage of data; it is a framework for using genome-scale experiments to perform predictive, hypothesis-driven science (H.-Y. Chuang et al. 2010). It is crucial that the models generated are consistent with literature and validated by detailed experimentation. The analysis of systematic data in fusion with system-wide measurements methods and computational approaches is used by system biology to formulate accurate hypothesis for further investigation and discovery.

Nowadays system biology comprises much of biology and other life sciences and is used to address many research problems. It has opened a new window for the study of several areas that some years ago seem difficult to understand and have affected the procedures of many others. The system approach in biology has a wide

range of applications that goes from biotechnology and genetic engineering to medical diagnosis and drug design and development. Even so, during the past few years economic forces and demands for a higher quality of life have repositioned research agendas. These agendas now include molecular diagnosis (Y. C. Wang & B. S. Chen 2011), systematic measurement and modeling of genetic interactions (G. Zhang et al. 2010), system biology of stem cells (Karantzali et al. 2010), and identification of disease genes (Z.-P. Liu et al. 2011).

### ***2.2.1 Bonding in Science***

The development of system biology has been facilitated through interdisciplinary collaboration. Many branches of science, if not most of them, have come together to give birth to this new division of biology. A vast number of recent technological advances have been made including quantitative high-throughput biological tools and mathematical software and algorithms that produce, archive, distribute, analyze, and model an ever-increasing amount of information. An advantageous combination between the feasibility of generating high quality data with the potential of new computer technology to analyze it has made possible the rapid incorporation of system biology into research practices.

Quantitative high-throughput biological tools; including, DNA sequencing, microarrays, and proteomics, implement global analysis and provide information about the state of the system and its dynamic behavior during perturbations. Proteomics is the characterization of the many proteins within a cell type. It is a powerful tool for identifying and quantifying large number of proteins. It also

analyzes protein modification (Ideker et al. 2001) through mass spectrometry (Krug et al. 2010) and NMR (Maldonado et al. 2010). DNA sequencing is used to analyze genomic DNA and cDNA and obtain information about genetic alterations, either simple sequence repeats or single nucleotide polymorphisms (SNP) (E. Y. Chan 2005). The analysis of genomic DNA is mainly used for genome wide scale sequencing purposes while cDNA analysis centers on the transcribed portion of the genome and is the foundation of most microarray technologies (Morozova & Marra 2008). High-throughput DNA sequencing is opening new fields and applications in biology and medicine beyond the genomic sequencing which was the original motivation. Together with cutting-edge technologies, such as, personal genomics analysis (Mirsaidov et al. 2010) and precise quantification of RNA for gene expression, DNA sequencing is modifying the currently study of biology (Ansorge 2009).

Researchers in system biology can be classified into two disciplines: computational biology, which develops tools and algorithms for system-level studies and network biology, which analyzes system properties using the tools and algorithms developed.

### ***2.2.2 Computational Biology***

The need of computational biology does not only rely on the increasing demand for handling huge quantities of experimental data generated by the different systematic techniques. The management and distribution of this information is just one side of this field of study. It also gives an answer to the intrinsic complexity of

biological systems (Kitano 2002a) by designing software and computational techniques for analyzing data and modeling biological systems and biological structures (Gerstein et al. 2007).

Significant resources have been allocated to many bioinformatics centers and projects around the world dedicated to data curation, its integration into databases and the development of tools necessary for its analysis. Projects like Mouse Genome Informatics (MGI) and the *Arabidopsis* Information Resource (TAIR) received in 2009 a yearly activity funding of 6.3 million US dollars and 1.6 million US dollars respectively. The same year the U.S. National Institute of Health (NIH) assigned 5% of its total budget of 20.9 billions US dollars to bioinformatics projects (Schofield et al. 2010).

Since the NIH issued its report in 1999 of the Biomedical Information Science and Technology Initiative (BISTI), the role of computation in system biology was categorically implemented (Friedman et al. 2004). During the last decade computational scientists in cooperation with biologists developed a vast number of computational tools that help in the effectively analysis of systematic data. This software meets a number of prerequisites in order to give an accurate approach of the system under study. First, the tools should be able to manipulate genome-scale data sets. Second, the tools should not have data type restrictions; they should be able to incorporate multiple measurements of the system. Third, they must work as a platform for mapping and modeling networks and pathways from the data available. Fourth, they should visually display the data, results, and models for further biological analysis (H.-Y. Chuang et al. 2010).

A good amount of software is now available for system biology research; some of it is intended for researchers with a solid background in computer science and another part is dedicated for biologists. Some of the bioinformatics tools available for biologists are: Cytoscape, NAViGaTOR, and VisANT for visualization, integration and analysis of biological networks, CellDesigner for schematization of gene regulatory networks and biochemical pathways, and PathwayAssists a mining tool for protein interaction and cellular pathway (H.-Y. Chuang et al. 2010). All of the previews bioinformatics packages are open source tools.

### ***2.2.3 Network Biology***

The best approach for analyzing biological systems is by mapping all of its components and determinate how the distinct architectural elements influence the behavior and function of the system. Network biology or network genomics, supported by the bioinformatics tools, incorporates the quantitative genome-scale data (e.g., gene expression and molecular interactions) to understand the function of genes in a systematic context (Hocquette 2005).

Vast information about the regulatory functions of individual genes is available, however, the final result of gene mutations or physical alterations is difficult to understand solely on this genomic information. One of the challenges in network biology is to investigate how each gene, through the interactions of a great number of factors, contributes to the final phenotype. The biological system as a whole is understood by analyzing the distinct cellular processes and mechanisms that

guaranty its correct function. Regulatory and signaling pathways are identified within the gene network that is constructed using molecular interactions information. The mapping of the biological mechanisms helps in the prediction of the role of these regulatory and signaling pathways in the general behavior of the system. This analytical work in combination with an exhaustive literature review is the point of departure for the generation of hypothetic-deductive models.

The construction of biological molecular interaction networks is crucial in the field of network biology. A rough estimate suggests that over 80% of the total proteins in humans do not operate alone but in complexes (Berggård et al. 2007). These protein complexes have major roles in the regulatory and signaling cellular pathways hence the study and understanding of them is an essential pillar in the behavioral research of the complete system. The first step towards the characterization and mapping of an interaction network involves the assembly of an exhaustive molecular interaction database that includes protein-protein interactions (PPI) and protein-DNA interactions. Other interactions between proteins and small molecules (lipids, drugs, metabolites or hormones) can be also included, however, it is difficult to perform large-scale measurement of these interactions and scientists do not often take them into consideration.

There are several approaches for identification of molecular interactions. For PPI the two most popular methods are the yeast two-hybrid system (Fields & Song 1989) and the protein co-immunoprecipitation (coIP) followed by mass spectrometry (Moresco et al. 2010). Recently new techniques have been introduced for a better characterization of PPI, especially in mammalian cells. For example, the mammalian two-hybrid system based on the functional tetracycline repressor (TetR) (Thibodeaux

et al. 2009) and the protein-fragment complementation assays (PCAs) (Michnick et al. 2007) are two of the new techniques used to investigate PPI, its dynamics, and its effects on any pathway of interest during specific perturbations. For protein-DNA interactions the most common procedure is chromatin immunoprecipitation (ChIP) generally used for identification of transcription factors and chromatin modifiers (Collas 2010).

Most of the interactions detected by any laboratory or research group are freely available in public databases (Razick et al. 2008), such as, DIP<sup>2</sup> (Salwinski et al. 2004), IntAct<sup>3</sup> (Aranda et al. 2010), MINT<sup>4</sup> (Chatr-Aryamontri et al. 2007), MIPS<sup>5</sup> (Pagel et al. 2005), BioGRID<sup>6</sup> (B. J. Breitkreutz et al. 2008), HPRD<sup>7</sup> (Keshava Prasad et al. 2009), and OPHID<sup>8</sup> (K. R. Brown & Jurisica 2005).

Researchers and scientists dedicated to network biology graphically represent interaction networks, identify regulatory pathways, map functional modules, and infer functional hypothesis. The role of network biology is to analyze the information available, display it in a schematic way and explain molecular functions within a network. Beyond that, system biology, as a junction of many scientific disciplines, models and predicts the behavior of complex systems and achieves a higher level of understanding.

---

<sup>2</sup> <http://dip.doe-mbi.ucla.edu>

<sup>3</sup> <http://www.ebi.ac.uk/intact>

<sup>4</sup> <http://mint.bio.uniroma2.it/mint/Welcome.do>

<sup>5</sup> <http://www.helmholtz-muenchen.de/en/ibis>

<sup>6</sup> <http://thebiogrid.org/>

<sup>7</sup> <http://www.hprd.org/>

<sup>8</sup> <http://ophid.utoronto.ca/>



## Chapter III : Previous Work

The present project began with a research study initiated at the Center for Renal Translation Medicine, Division of Nephrology and Hypertension, Department of Medicine, University of California, San Diego, by Dr. Timothy Ravasi, Dr. Satish P. RamachandraRao, and Dr. Kumar Sharma. The overall objective of that proposal was: 1) to identify potential biomarkers in diabetes and in its major long-term complication diabetic nephropathy (DN) and 2) to determine the mechanism of action of pirfenidone (PFD), an antifibrotic drug, in the arrest of the progression of diabetic nephropathy after manifestation.

Pirfenidone (5-methyl-1-phenyl-2-(1H)-pyridone) (Figure 3) is a pyridine molecule that is able to cross the cell membrane without the assistance of a receptor and as a result is easily absorbed in the gastrointestinal tract. It can reach most tissues and travel across the blood-brain barrier. It does not show significant toxicity at doses up to 2500 mg/day and is almost fully eliminated through the urine 6 hours after ingestion (Macías-Barragán et al. 2010). PFD was first developed as an analgesic, antipyretic, and anti-inflammatory drug. Later on, based on the hypothesis that anti-inflammatory drugs might have antifibrotic properties, PFD was found to be useful for the treatment of fibrosis in the hamster pulmonary fibrosis model (Gan et al. 2011). PFD was first clinically used in idiopathic pulmonary fibrosis patients (Raghu et al. 1999) and has been tested as an antifibrotic drug in other tissues, mainly liver (Macías-Barragán et al. 2010) and kidney (Djamali & Samaniego 2009). PFD acts by reducing the production of transforming growth factor  $\beta 1$  (TGF- $\beta 1$ ) which is

responsible for stimulating fibronectin synthesis and reducing fibroblast proliferation and collagen production from fibroblast (Gan et al. 2011); however, the antifibrotic properties and the mechanism of action of PFD are poorly understood (Shihab 2007).

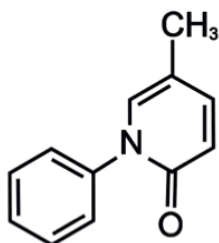


Figure 3  
Chemical structure of PFD

The following results are part of the findings from the previous experiments done by RamachandraRao, S. *et al.* (RamachandraRao et al. 2010; RamachandraRao et al. 2009) regarding the effect of PFD in the treatment of diabetes. For more details please refer directly to the publications.

The first section of experiment consisted in the treatment of murine mesangial cells (MMC) with PFD. Mesangial cells (MC) are connective tissue cells that surround the filtration capillaries within the glomerulus (Stockand & Sansom 1998). Some of the physiological roles of MC include: the secretion and signaling of hormones and growth factors, the synthesis of ECM in order to provide structure and stability to the filtration barrier, and the regulation of the glomerular filtration rate (GFR) (Stockand & Sansom 1997). The pathological role of MC in the development of DN arose from the crucial position of MC in glomerular capillaries in addition to their ability to produce ECM proteins and to regulate the GFR. DN is characterized by an accumulation of extracellular matrix (ECM) proteins in the glomerular mesangium, furthermore, an increase in the GFR is strongly related to the

development of DN. These anomalies could be caused by functional variations of the glomeruli in diabetic patients, specifically, in the MC (Haneda et al. 2003).

The treatment (MMC) line with PFD reduces the TGF- $\beta$ 1 promoter activity and the total secretion of TGF- $\beta$ 1 protein. It also blocks the TGF- $\beta$ 1 signaling measured in terms of Smad2 phosphorylation. As a response to TGF- $\beta$ 1 signal the Smad2 protein is phosphorylated (Poncelet et al. 1999). The TGF- $\beta$ 1-induced Smad2 phosphorylation is inhibited with PFD treatment as well as the total production of Smad2 protein. TGF- $\beta$ 1 is a strong profibrogenic factor it regulates the maintenance and development of fibrogenesis and it is responsible for keeping the balance between the synthesis and degradation of the ECM (Cheng et al. 2009). New therapeutic treatments for fibrotic diseases focused on blocking the protein production and signaling pathway of TGF- $\beta$ 1 to prevent the synthesis of ECM and accelerate its degradation (Hsu et al. 2010). The mesangial cells also showed a reduction in the stimulation of TGF- $\beta$ 1 on the mRNA levels of type I and type IV collagen after treatment with PFD. The functional changes in glomeruli and MC are considered to be caused by abnormalities in the glucose-induced signaling pathway (Figure 4) of MC that results in the pathogenesis of DN. The capacity of PFD to inhibit the TGF- $\beta$ 1 promoter activity and the production of the protein reduces the glucose uptake in mesangial cells even under high-glucose conditions. This could hinder the production of collagen and the synthesis of ECM.

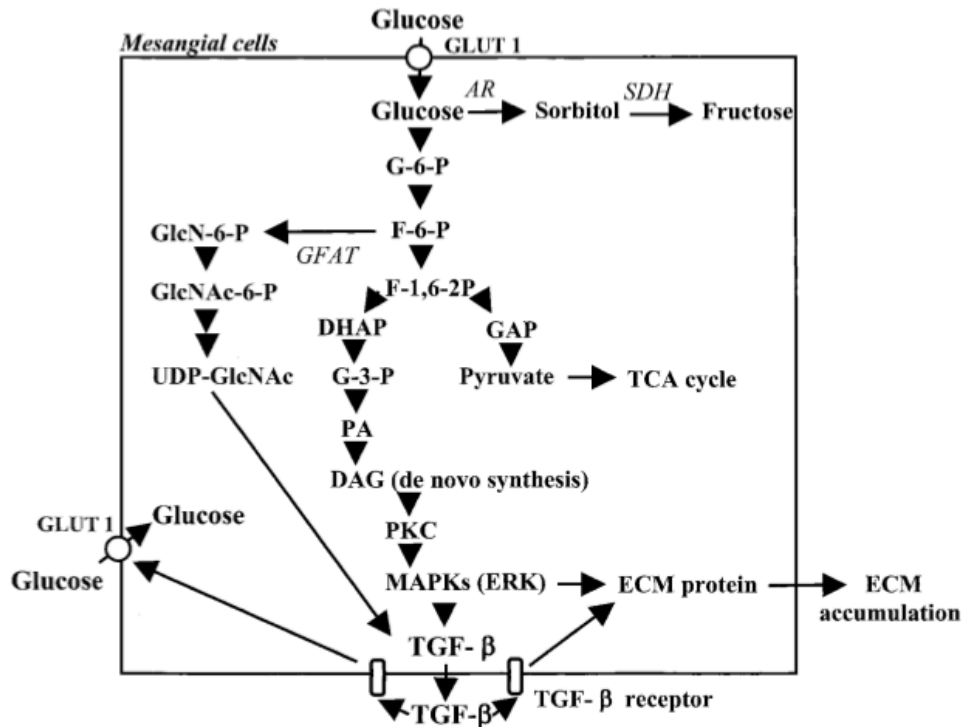


Figure 4  
Glucose-induced signaling pathway in mesangial cells  
(Haneda et al. 2003)

The other section of the experiment consisted in the treatment of murine models with the drug. The model selected for this purpose was the *db/db* mouse. This murine model is well characterized and has been subject to intensively investigation. The model exhibits a consistent increase in albuminuria and mesangial matrix expansion making it an ideal model for the investigation of progressive diabetic renal disease, such as, DN (Sharma et al. 2003). The similarities of this model with human DN, which include, a closely mimic of the progressive nature of mesangial matrix expansion, renal hypertrophy, albuminuria, and glomerular enlargement have position the *db/db* mouse as the preferred model to investigate the role of several pathways in the development of diabetic renal disease and the use of drugs to halt its progression.

On the *db/db* mouse model, four weeks of PFD treatment reduced the mesangial matrix expansion and the gene expression of renal type I collagen, type IV collagen, and fibronectin, however, the blood glucose levels and albuminuria was not affected.

The reduction of mesangial matrix expansion is crucial in the treatment of DN. The mesangial matrix expands as a result of the accumulation of proteins normally present in the mesangial matrix. The composition of the mesangial matrix is altered, in diabetic patients, mainly by the presence of high levels of glucose (Abrass 1994). Normal mesangial matrix is mainly composed by collagen and fibronectin, elevated glucose levels increase protein synthesis and protein accumulation. During long periods of uncontrolled glucose levels the synthesis of mesangial matrix increases and the composition of the matrix changes, these abnormalities lead to the occlusion of the glomerular capillaries and to renal failure.

Proteomics of kidney from the nondiabetic, diabetic, and diabetic treated with PFD mice was performed using liquid chromatography with subsequent tandem mass spectrometry (LC-MS/MS). Twenty-one unique mouse proteins were found in the PFD-treated diabetic kidney. These unique proteins were analyzed with respect to their physical interactions in a PPI network based on human PPI information due to the lack of murine data. Out of the 21 unique proteins, 14 have human orthologs and 11 could be mapped into the PPI network. The result was a network centered in the PFD-treated unique proteins with a total of 518 proteins and 655 interactions with a significantly enriched biological function in posttranslational or posttranscriptional regulation and mRNA processing.

The results indicate that PFD inhibits TGF- $\beta$ 1 promotion activity and protein secretion as well as TGF- $\beta$ 1 signaling through repression of TGF- $\beta$ 1-induced Smad2 phosphorylation, although the exact mechanism of action is unclear. The proteomic and bioinformatics study suggest that mRNA processing and translation are affected by PFD arising the hypothesis that the consequences of PFD in the diabetic kidney are mediated via mRNA translation pathway.

Complex diseases, such as diabetes, are difficult to study because there are many factors acting together that contribute to their onset, development, and subsequent long-term complications. A classic biological approach is limited in the examination of complex diseases since the interactions and relations between the distinct elements are not simple and the effect of perturbations is not homogenous. Hence, the analysis of complex diseases should be addressed from a system biology point of view. The present project with a complete systematic approach acts as a complementary investigation to the previous work. It analyzes at an expression level the repercussion of DM and the mechanism of action of PFD in the treatment of the disease. The contribution of this analysis will provide new clues for the construction of models and hypothesis and will give a deeper insight into the behavior of the disease.

## Chapter IV : Results and Discussion

### 5.1 Preliminary Data Treatment

The raw data obtained from the microarray experiment consisted of a total of 45,284 probes. Some of the probes represent sequenced cDNA clones that are likely to be new mouse genes but currently remain classified as “hypothetical proteins”, “unclassifiable transcript”, or “motif-containing protein” (The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium 2001). These probes do not have specific gene information, making them useless for the present analysis. Other probes are duplicated across the microarray chip resulting in a double quantification of the same gene. The prefiltering of the microarray data comprises the removal of all the probes lacking specific gene information and the calculation of the arithmetic mean of the intensity measurements for each of the probes with two or more sets of data.

The hypothesis behind microarray assays that the measured intensity represents the expression level or the mRNA concentration of the gene is true for most cases. Hybridization experiments such as the one performed by Lockhart *et al.* (1996) described a linear correlation (Figure 5) between the concentration of RNA and the hybridization intensity. However there are multiple factors, such as, the size of the target DNA, the probe density, and the GC content of the sequence that affect the reaction-diffusion process in the heterogeneous DNA hybridization. These factors alter the linear correlation and contradict the hypothesis.

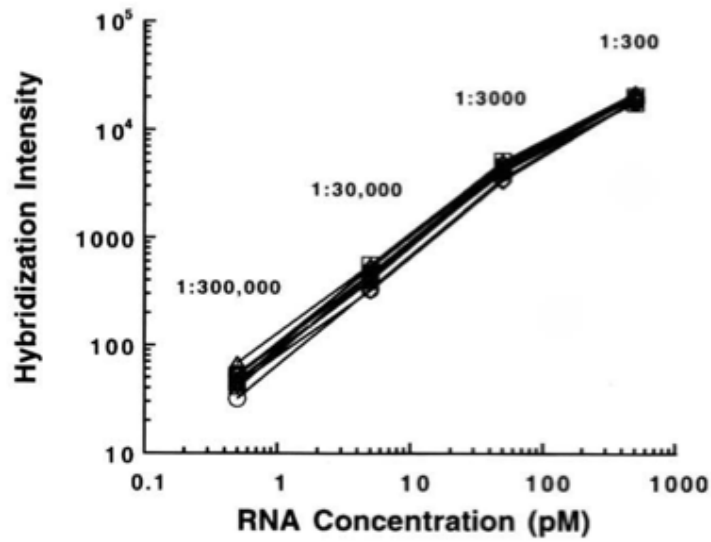


Figure 5  
Intensity vs. RNA concentration correlation  
(D. J. Lockhart et al. 1996)

Chan *et al.* (1995) showed that the hybridization reaction rate represented by the parameter  $f_{2D}$  (two-dimensional fraction) and measured as a ratio of the hybridization rate due to lateral diffusion to the total hybridization rate is severely influenced by the size of the target DNA (Figure 6). For a fixed distance between immobilized probes ( $R_b$ ) the reaction rate drops drastically while the size of the oligonucleotides gets bigger. The graph shows that the targets with a bigger base number will perform a slower hybridization reaction.



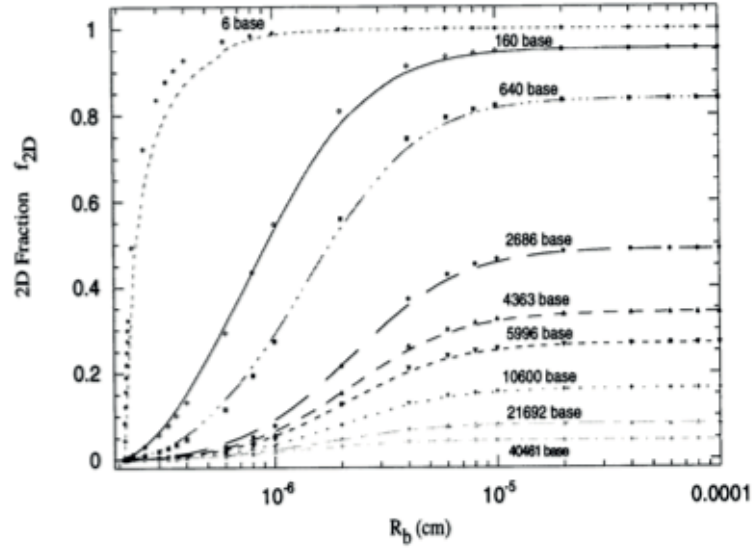


Figure 6  
 $f_{2D}$  as a function of inter-probe spacing ( $R_b$ ) for various target size  
 (V. Chan et al. 1995)

Other experiments indicate that the GC content of either the probe or the target has a great influence in the final measurement of the spots intensity. The findings of Livshits and Mirzabekov (1996) indicated that the recorded intensity of two spots one with a strongly binding GC-rich oligonucleotide ( $S$ ) and the other with a weakly binding GC-poor oligonucleotide ( $W$ ) have clearly different fluorescent intensities (Figure 7). Statistical and thermodynamic methods indicate that the free energy of short sequences decreases while the GC content of the oligonucleotide increases (Tulpan et al. 2010). Meaning that for GC-rich sequences the hybridization reaction is favored and the duplex formed is more stable.

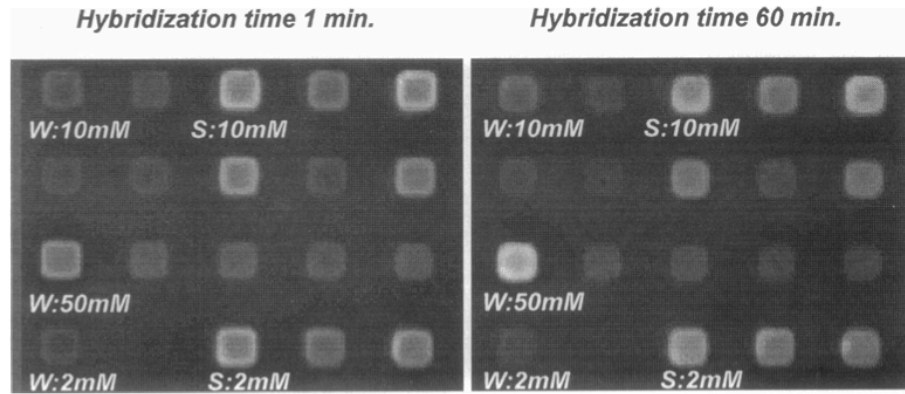


Figure 7  
Hybridization of fluoresceine-labeled oligonucleotide on a microchip  
(Livshits & Mirzabekov 1996)

The different factors that impact the performance of a microarray experiment are difficult to control. The data obtained from microarrays assays is subject to systematic variations caused by many sources. Most microarray experiments are not replicable and there is not a standardized guideline to produce uniformly reliable results (Reimers 2005). In order to minimize the influence of these fluctuations in the expression analysis of the diabetic mice it was decided to segregate from the data all the probes that do not have an intensity measurement above 100 in any of the samples. The cutoff value was selected based on empiric knowledge. The probes with an intensity value below 100 usually have reaction rate limitations and the intensity is not a reliable measurement of expression levels.

After removing the probes without gene information, averaging the probes with multiple measurements, and separating the probes with reaction rate limitation, the final dataset was obtained. The data employed for the expression-based analysis comprises 5,589 unique probes with specific gene information.

## 5.2 Statistical Analysis of Microarray

The statistical analysis in an expression-based study identifies the genes that are differentially regulated across the different experimental groups. This analysis classified the genes into expression clusters. The expression clusters for the present work are defined as groups of genes that are statistically significant expressed in an experimental group or set of experimental groups. For the purpose of the present work genes were classified in one of the following categories or expression clusters, represented by the Venn diagram in Figure 8:

- a) Control
- b) Diabetic (*db/db*)
- c) Treated (*db/db* + PFD)
- d) Control and Diabetic
- e) Control and Treated
- f) Diabetic and Treated
- g) All Groups

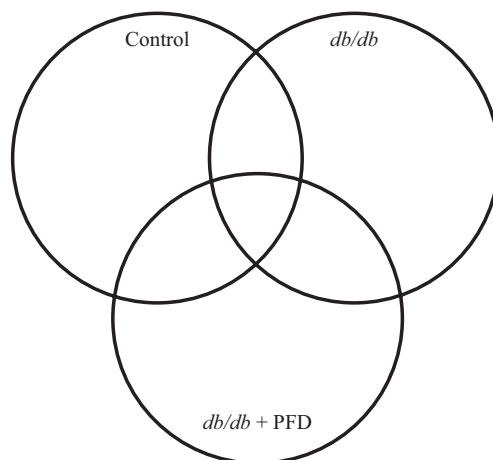


Figure 8  
Venn diagram of the expression clusters

The statistical analysis of the data was done using the statistical software SAM (Significance Analysis of Microarray). SAM classified the genes as potential significant genes based on a robust permutation method and relies on the false discovery rate (FDR) and  $q$ -values to measure how significant a gene is. The experimental groups were analyzed in a pairwise basis and the datasets were treated as *Two Class (unpaired)* data. The results of each analysis were compared in order to eliminate duplicated results and classify the genes in a particular expression cluster. For the case of All Groups Expression Cluster the three experimental groups were treated together as *Multiclass* data. All the statistical tests were ran using a FDR of 10%. Even though there is no convention about the ideal FDR, a rate of 5 or 10% is normally acceptable (Benjamini & Hochberg 1995; Reimers 2005).

The results of the statistical tests (Figure 9) showed 873 genes significant expressed in the control experimental group, 182 in the diabetic group, and 118 in the treated group. There are 170 genes significant regulated in the control and diabetic group, 195 in the control and treated group, and 1055 in the diabetic and treated group. A total of 304 are classified as significant expressed genes in all the groups.

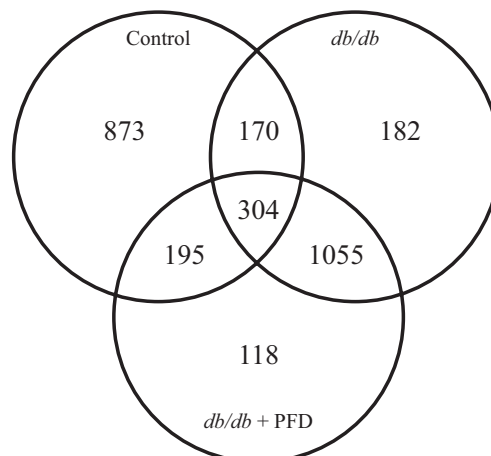


Figure 9  
*Mus musculus* gene expression clusters

After the data set was statistical analyzed the outcome was 7 expression cluster that integrate the genes that show a significant expression in a particular experimental group or a set of experimental groups. For the purposes of the present analysis the expression clusters: control, diabetic, treated, diabetic and treated, and control and treated are of great relevance (Figure 10). The control expression cluster works as a reference framework. The diabetic expression cluster represents genes that show major changes during the disease while the treated expression cluster represents genes with significant regulation in the presence of the drug. The genes significantly annotated in the diabetic and treated expression cluster are dysregulated in the diabetic state and their expression is not affected by the presence of the drug, in some cases the expression dysregulation increases after treatment. Finally, the control and treated expression cluster gathers genes that return to normal expression levels after drug treatment, this group is of particular interest to understand the mechanism of action of PFD. The discussion will be centered in these 5 expression clusters.

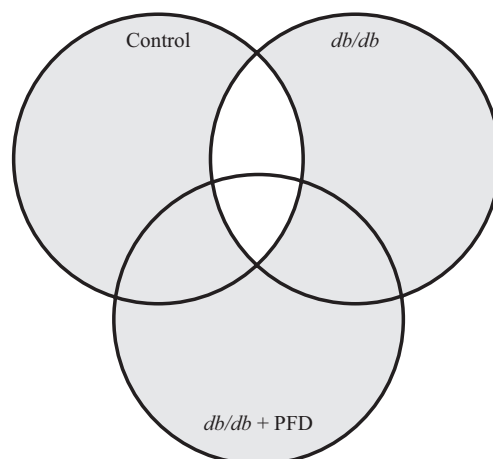


Figure 10  
Relevant expression clusters

### 5.3 Human-Mouse Orthology

The mapping of the genes onto a PPI network is essential in the study of complex diseases, such as diabetes. The NIA Mouse Protein-Protein Interaction Database<sup>9</sup> contains a vast dataset of interactions between proteins in *Mus musculus*. However the interactions within this database were predicted on the basis of their orthologous information in other organisms and as a consequence the interactions are subject to false positive interactions and false positive interologous (Yellaboina et al. 2008). Some studies suggest that only about 16 to 30% of the protein interactions within an organism is transferable to another one despite a correct orthologous matching of the protein (Mika & Rost 2006). The experimental identification of proteins in *M. musculus* is limited and the implementation of predicted PPI information may cause variations in the analysis of the data. In the other hand, the interactions between proteins in *Homo sapiens* have been experimental recorded with high-throughput experiments and most of the interactions have been confirmed with detailed biochemical experimentation (Rual et al. 2005). Although the human PPI dataset is not exempt from experimental bias the experimental verification of the interactions makes it a more reliable source. Taking this in consideration, it was decided to map the genes onto a human PPI network.

Orthology analysis is the identification of orthologous genes across species and it is a common approach in comparative genomics. Orthology connects genomes from different organisms and allows the exchange of gene annotations between them (Sennblad & Lagergren 2009). The hypothesis behind this analysis is that

---

<sup>9</sup> <http://lgsun.grc.nia.nih.gov/mppi/index.html>

orthologous genes descend from a common ancestor and are likely to perform the same function (Ostlund et al. 2010). There are a great amount of algorithms used for the prediction of orthologous genes. These algorithms identify orthologous either by best BLAST hit or phylogenetic tree reconstruction. Either way none of the prediction orthologous methods are free of errors and can result in the prediction of false positive orthologous. Nevertheless, orthology between human and model organism has been a useful tool in the detection of human disease genes and in drug discovery (Doyle et al. 2010).

The human orthologous counterparts of the mouse genes within the expression clusters were identified with the help of the web-based software BioMart. BioMart matches gene orthologous based on the Ensembl ortholog/paralog prediction pipeline that is a combination of best BLAST hit and maximum-likelihood phylogenetic (Hubbard et al. 2007). In every expression cluster more that 90% of the mouse genes have a human orthologous (Figure 11).

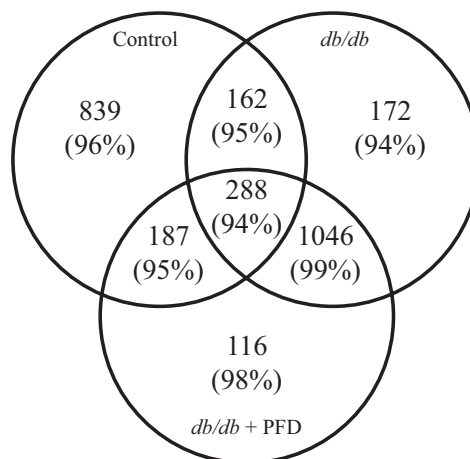


Figure 11  
*Homo sapiens* orthologous gene expression clusters

## 5.4 Gene Ontology Enrichment Analysis

An enrichment analysis was performed to each of the expression clusters using the human orthologous genes. This analysis identifies biological process and pathways that are overrepresented in a given set of genes in comparison to a background or a reference gene list (Lammers et al. 2010). The background is usually the complete genome of the specie under study. The biological process, cellular components, and molecular functions are represented by gene ontology (GO)<sup>10</sup> terms while the biological pathways are identified by the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>11</sup> annotations.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used for the enrichment analysis of each individual expression cluster. The genes that conform the complete human PPI network are used as background. All the biological process and pathways listed in Table 3 are significant as indicated by the *p*-value. Usually the terms with a *p*-value < 0.05 are considered significant (Lammers et al. 2010). There are terms that have a *p*-value > 0.05 but were included in the analysis (e.g., the terms in the diabetic and control group: inflammatory response (GO:0006954) and positive regulation of macromolecule and cellular biosynthetic process (GO:0009891) with *p*-value of  $1.27 \times 10^{-02}$  and  $2.13 \times 10^{-02}$  respectively). These annotations were consider significant since they have fold enrichment (F. E.) > 2.

---

<sup>10</sup> <http://www.geneontology.org/>

<sup>11</sup> <http://www.genome.jp/kegg/>



Table 3  
Enrichment analysis of the expression clusters

Term	<i>p</i> -value	F. E. *
<b>Control Group</b>		
GO: 0051276~chromosome modification and organization	$3.98 \times 10^{-04}$	1.97
GO:0044093~positive regulation of molecular function	$1.34 \times 10^{-03}$	1.62
GO:0010033~response to organic substance	$4.81 \times 10^{-03}$	1.45
<b>Diabetic Group</b>		
GO:0051329~interphase of mitotic cell cycle	$6.09 \times 10^{-05}$	8.90
KEGG:04115~p53 signaling pathway	$7.86 \times 10^{-04}$	7.09
GO:0033554~cellular response to stress	$4.62 \times 10^{-03}$	2.35
GO:0000082~G1/S transition of mitotic cell cycle	$5.22 \times 10^{-03}$	10.17
GO:0032582~negative regulation of gene-specific transcription	$7.56 \times 10^{-03}$	9.04
<b>Treated Group</b>		
GO:0007389~pattern specification process	$2.99 \times 10^{-04}$	5.77
GO:0048729~tissue morphogenesis	$5.84 \times 10^{-03}$	4.95
<b>Control and Diabetic Group</b>		
GO:0006954~inflammatory response	$1.27 \times 10^{-02}$	4.12
GO:0009891~positive regulation of macromolecule and cellular biosynthetic process	$2.13 \times 10^{-02}$	2.36
<b>Control and Treated Group</b>		
GO:0007243~protein kinase cascade	$3.92 \times 10^{-03}$	3.50
GO:0006396~RNA processing	$4.28 \times 10^{-03}$	2.89
KEGG:04010~MAPK signaling pathway	$3.82 \times 10^{-02}$	3.09
<b>Diabetic and Treated Group</b>		
GO:0043933~cellular macromolecular complex subunit organization	$2.58 \times 10^{-04}$	1.47
GO:0006091~generation of precursor metabolites and energy	$5.75 \times 10^{-04}$	1.65
GO:0015031~protein transport	$2.19 \times 10^{-03}$	1.34
GO:0048545~response to steroid hormone stimulus	$4.71 \times 10^{-03}$	1.87
GO:0044271~nitrogen compound biosynthetic process	$5.00 \times 10^{-03}$	1.56
GO:0032268~regulation of cellular protein metabolic process	$7.48 \times 10^{-03}$	1.44
KEGG:00190~oxidative phosphorylation	$8.67 \times 10^{-03}$	1.66
KEGG:00982~drug metabolism	$9.06 \times 10^{-03}$	2.21
KEGG:05211~renal cell carcinoma	$1.00 \times 10^{-02}$	2.09

\* Fold Enrichment

The enrichment analysis of the expression cluster provides a first insight of the regulatory pathways and biological functions that are affected in each experimental condition. In the diabetic expression cluster there is a particular enrichment in the gene ontology terms: cellular response to stress ( $p\text{-value} = 4.62 \times 10^{-03}$ ) and negative regulation to gene-specific transcription ( $p\text{-value} = 7.56 \times 10^{-03}$ ), in addition to the KEGG pathway: p53 signaling pathway ( $p\text{-value} = 7.86 \times 10^{-04}$ ). These 3 terms are of particular interest since the dysregulation of genes involved in these biological functions can be a response to the stress conditions present in diabetes. An increase concentration of free fatty acids promotes the mitochondrial  $\beta$ -oxidation mechanism, producing a great amount of reactive oxygen species (ROS), which induces insulin resistance (Y. C. Chang et al. 2010). The negative regulation of gene-specific transcription could be also a response to the oxidative stress. The endoplasmic reticulum (ER) stress is promoted by a high concentration of ROS, leading to the activation of the unfolded protein response (UPR). The UPR enhance the protein folding capacity of the ER and inhibits the production of new proteins (Shameli et al. 2007). ER stress is associated with  $\beta$ -cell apoptosis and a potential contribution to the pathogenesis of diabetes (Oyadomari et al. 2002).

The enrichment analysis of the control and treated expression clusters show enrichment in RNA processing ( $p\text{-value} = 4.28 \times 10^{-03}$ ). This concurs with the research studies of RamachandraRao, *et al.* (2010) that propose an mRNA translation activity of PFD. Their findings suggest that the beneficial effects of PFD are mediated via mRNA translation pathway (RamachandraRao et al. 2010). The gene ontology enrichment of the control and treated expression cluster in RNA processing indicates that genes involve in this biological function show similar expression levels in the

control as well as in the treated experimental group. PFD seems to be compensating the dysregulation of genes involved in the RNA processing mechanism in the diabetic state by returning the expression of these genes to their normal levels. This is confirmed by looking at the expression behavior of the genes significant expressed in the control and treated group that conform the mention ontology group (Figure 12).

The diabetic and treated expression cluster is the biggest cluster with more than 1000 genes. This is expected since most of the genes are not affected by the presence of the drug and its expression will remain the same in the treated group as in comparison with the diabetic group. In some cases the dysregulation of the genes get worse in the treated group probably to the terminal condition of the disease. This expression cluster is enriched in generation of precursor metabolites and energy ( $p\text{-value} = 5.75 \times 10^{-04}$ ) and response to steroid hormone stimulus ( $p\text{-value} = 4.71 \times 10^{-03}$ ), two terms tightly related to the functional activity of mitochondria. Moreover, enrichment in the KEGG pathway: oxidative phosphorylation ( $p\text{-value} = 8.67 \times 10^{-03}$ ), the main source of energy production, suggests that the disease is affecting the energy production of the cell and this alteration cannot be restored after drug treatment.

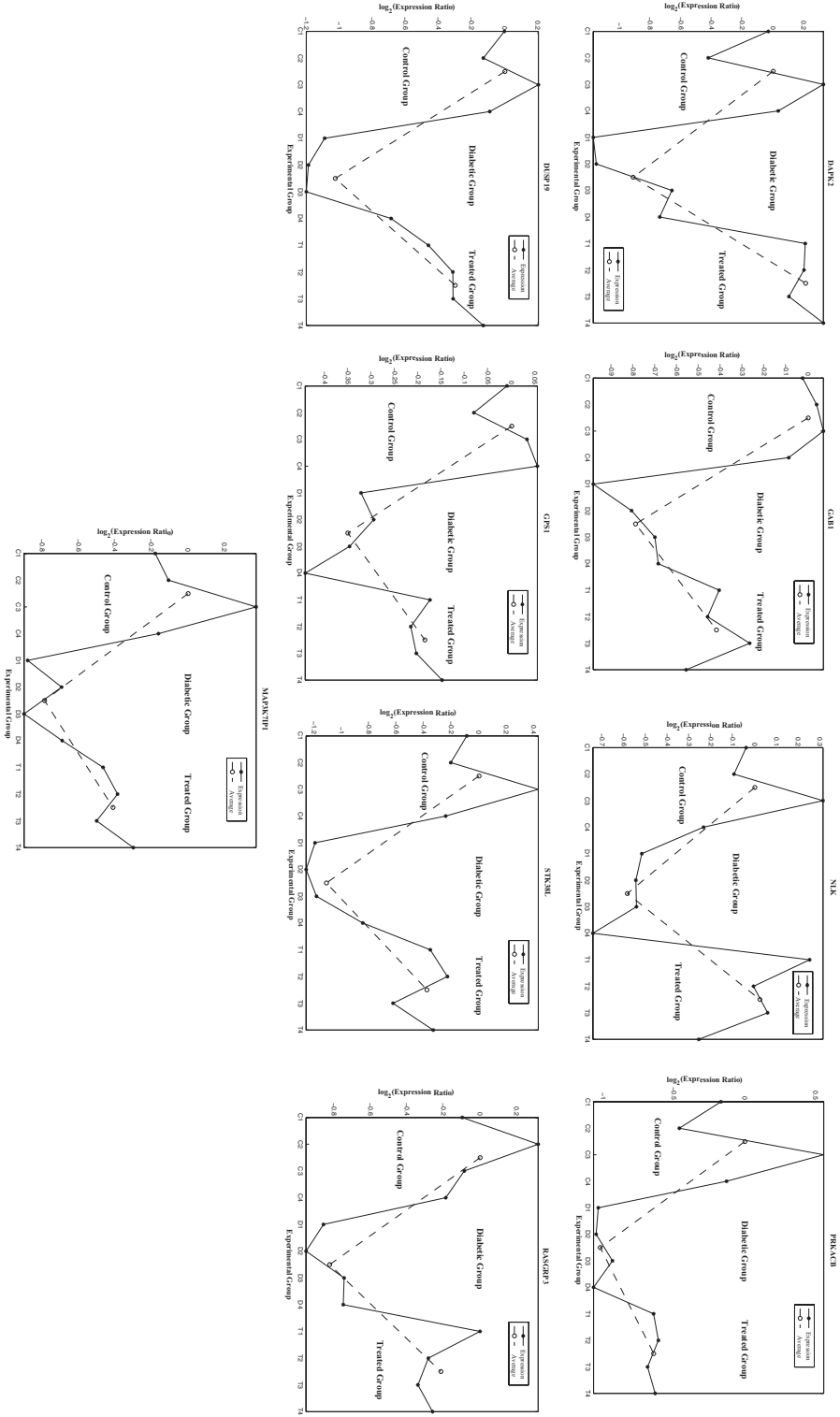


Figure 12  
Expression profile of genes significant expressed in the control and treated expression cluster annotated in RNA processing

## 5.5 Protein-Protein Interaction (PPI) Network

The genes that compose the expression clusters were mapped onto a human PPI network in order to build the corresponding PPI subnetwork for each of the expression clusters (Figure 13 - Figure 19). The complete human network consists of more than 58,000 interactions among 11,262 proteins. The genes annotated within a significant term were highlighted in its corresponding PPI subnetwork. The subnetworks were mapped and visualized using Cytoscape as a platform. The top 6 hubs of the complete PPI network (Table 4), all of them with a degree  $k < 200$ , were mapped along the significant expressed genes of each group.

Table 4  
Major network hubs

Entrez GeneID	Symbol	$k$
3172	HNF4A	1957
3175	ONECUT1	272
6927	HNF1A	266
7157	TP53	249
5432	POLR2C	203
5430	POLR2A	202

The HNF4A gene is significantly regulated in the control experimental group (Figure 13) while TP53 is regulated in the control and treated group (Figure 17). The other hubs are not annotated as significant genes for any of the expression clusters. Together these 6 major hubs interact with 2490 proteins, which represent 22% of all the proteins within the complete PPI network. They were included on the subnetworks since they have a key structural role and keep the subnetworks connected. The major hubs were not included into the control subnetwork (Figure 13)

and the diabetic and treated subnetwork (Figure 19). These two subnetworks are already highly connected and do not required intermediary genes that link the different components of the network. The areas highlighted with thicker edges in Figure 13 and Figure 19 are zones of the subnetwork that are very well connected.

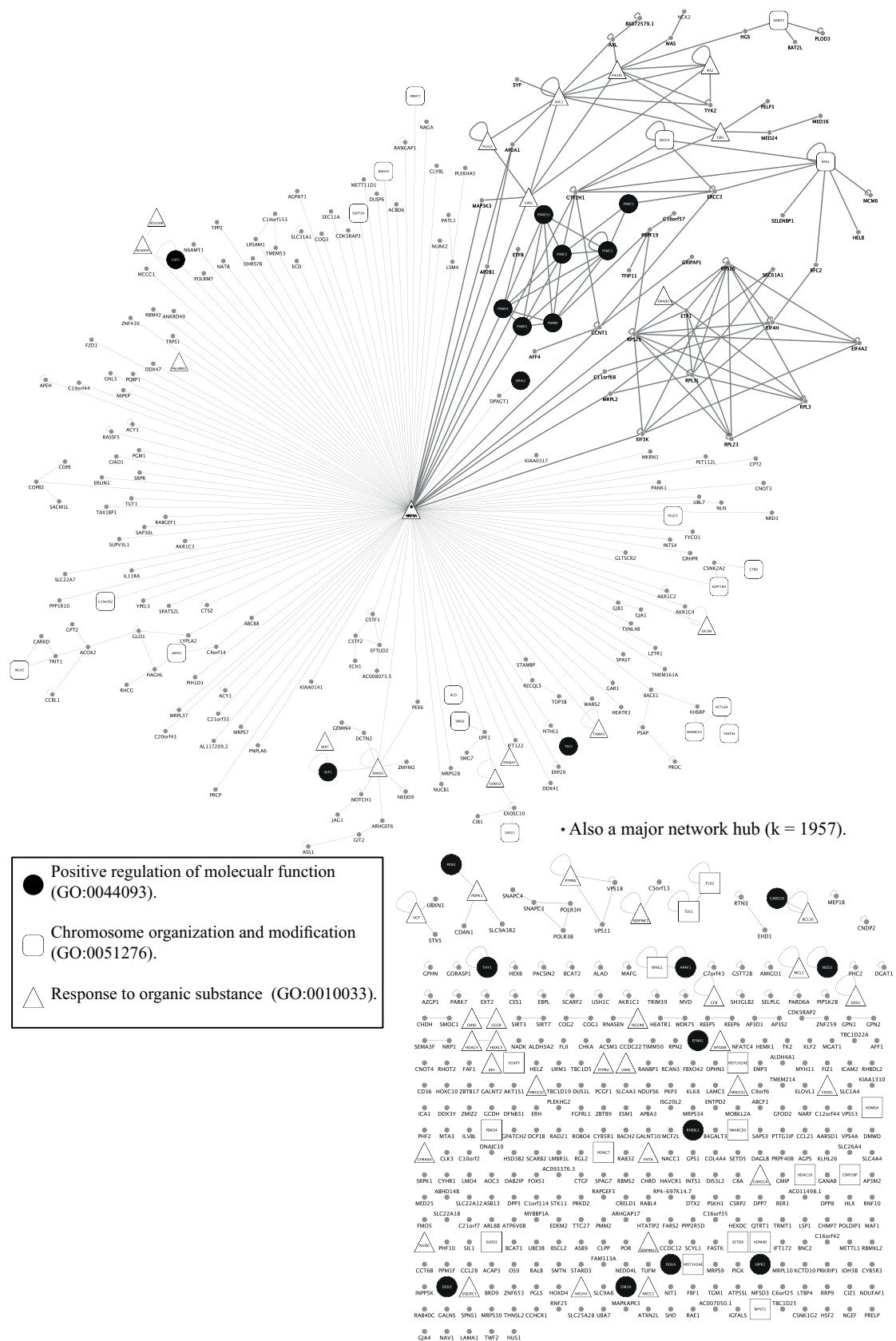


Figure 13  
Control PPI subnetwork

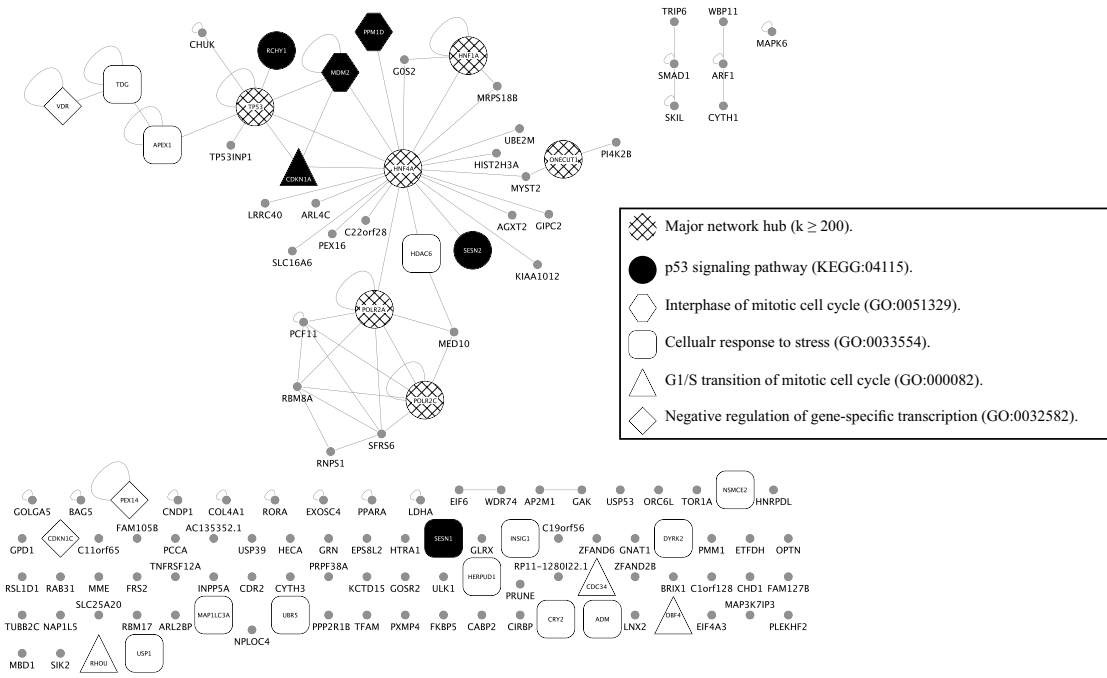


Figure 14  
Diabetic PPI subnetwork

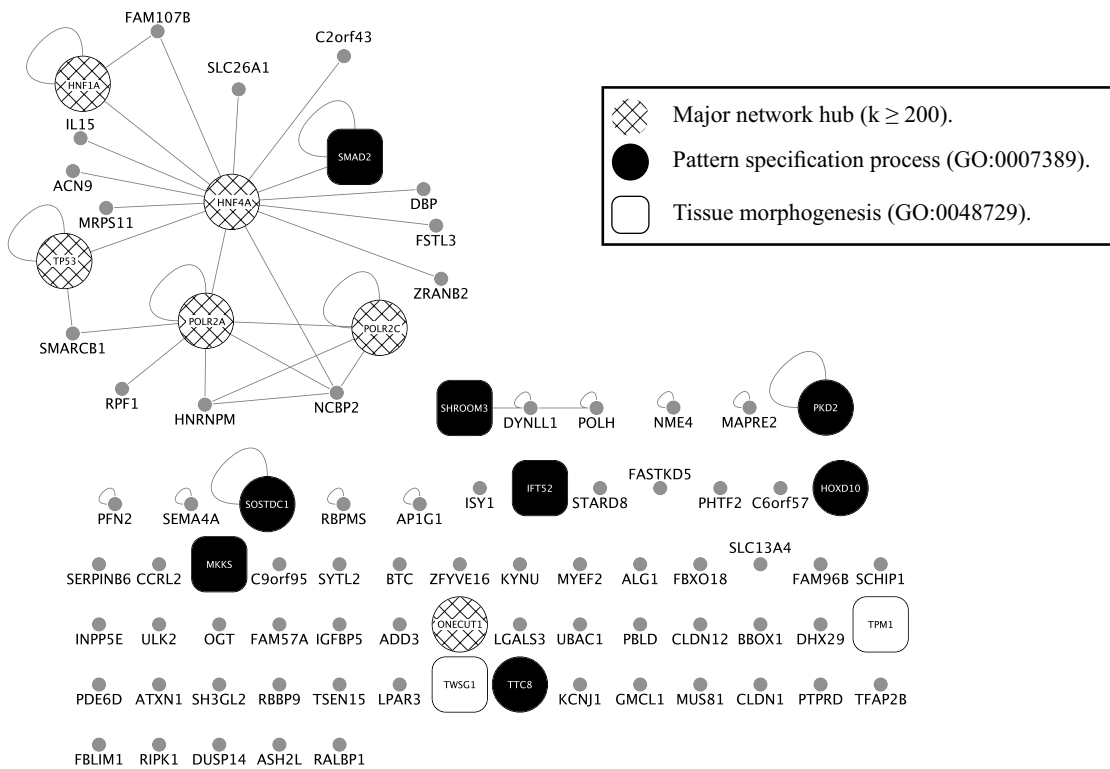


Figure 15  
Treated PPI subnetwork



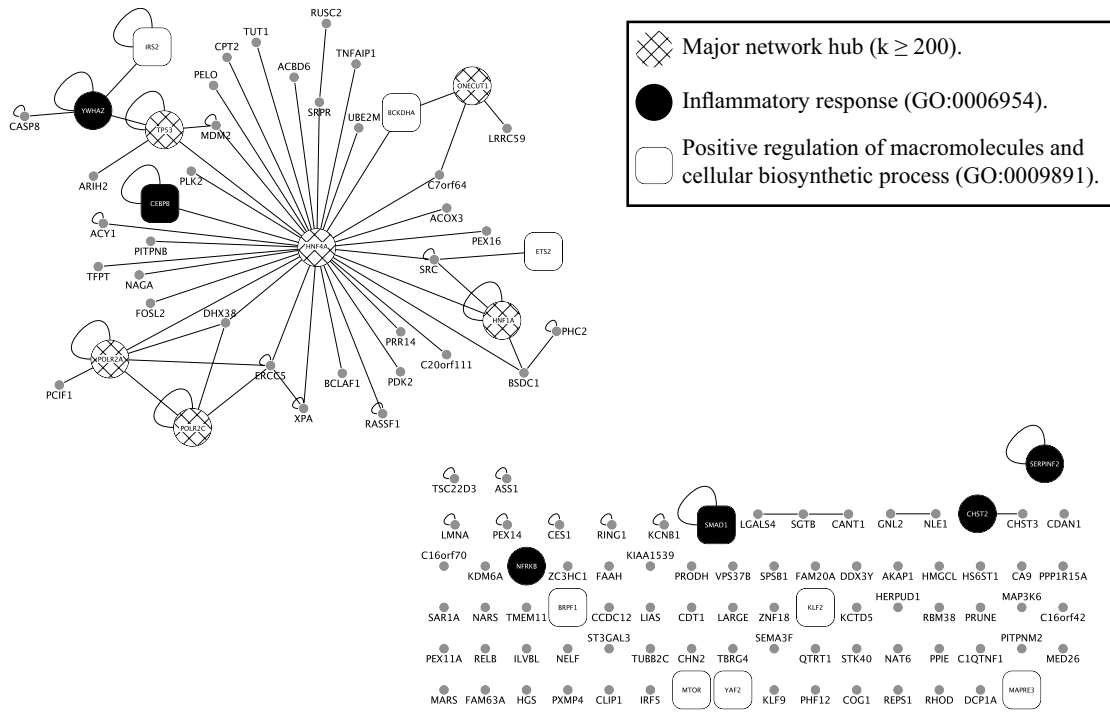


Figure 16  
Control and Diabetic PPI subnetwork

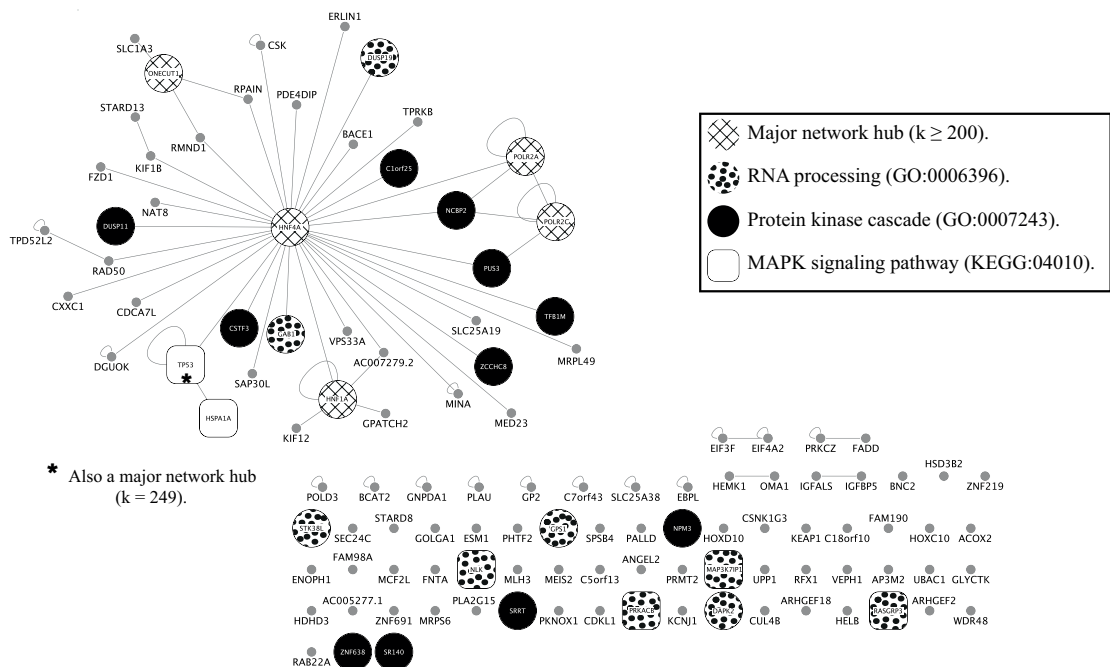


Figure 17  
Control and Treated PPI subnetwork

A particular subnetwork within the diabetic and treated PPI subnetwork requires further analysis. The subnetwork (Figure 18) is integrated by genes enriched in regulation of cellular protein metabolic process ( $p\text{-value} = 7.48 \times 10^{-03}$ ), in specific; this group of genes (Table 5) is part of the proteasome complex ( $p\text{-value} = 2.4 \times 10^{-11}$ ). The disruption of the ubiquitin-proteasome pathway can lead to an abnormal or accelerated degradation of proteins that will result in the pathogenesis of many diseases. The ubiquitin-proteasome system may play an important role in the onset of diabetes by promoting the degradation of insulin receptor substrate (IRS)-1 and 2 that leads to inhibition of insulin pathway (Balasubramanyam et al. 2005).

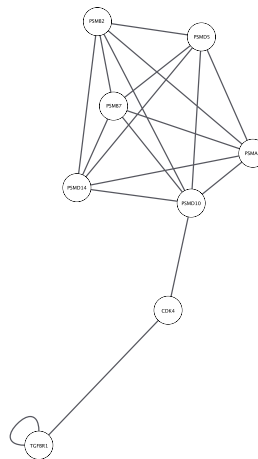


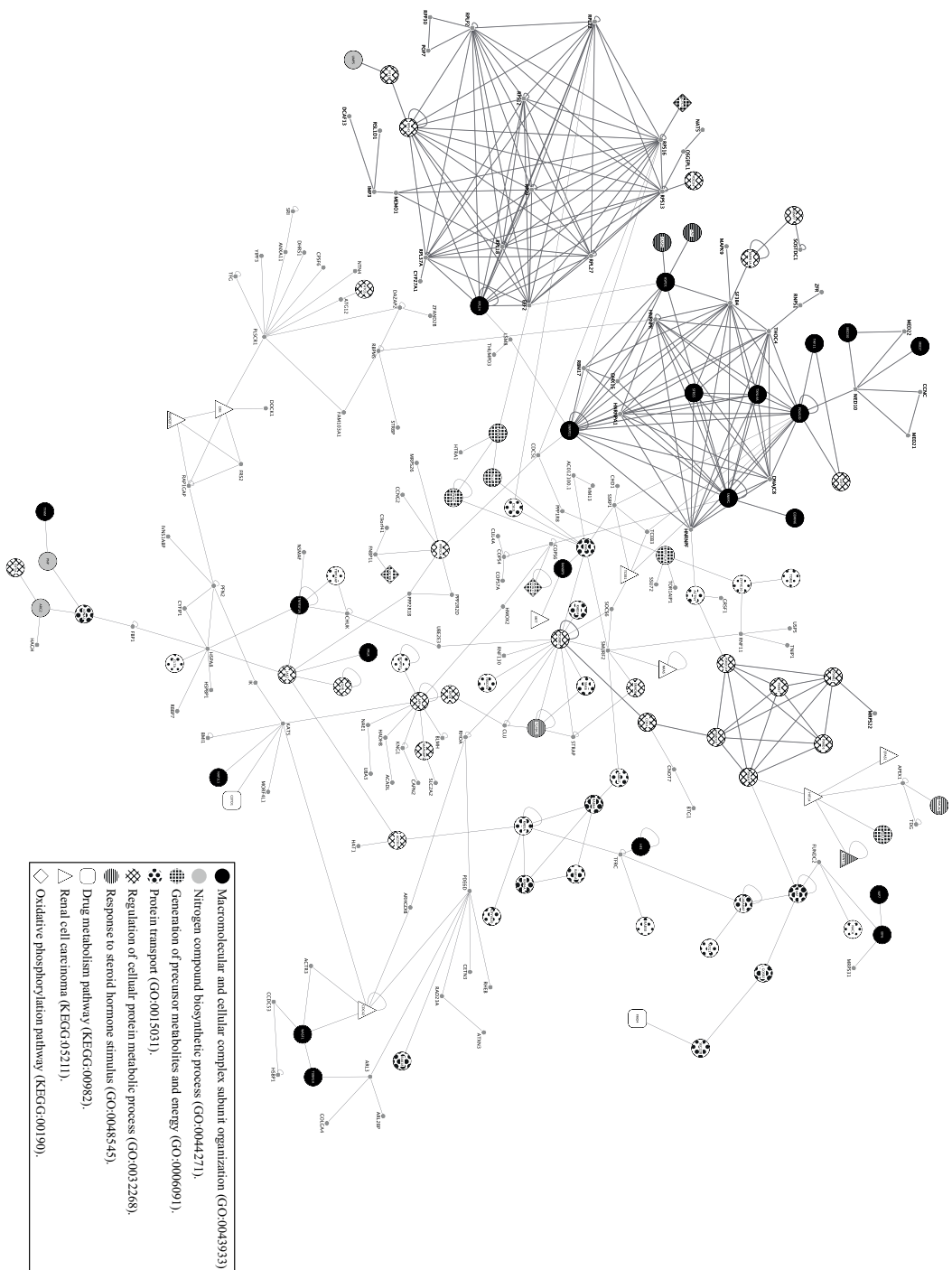
Figure 18

PPI subnetwork enriched in protein metabolic complex part of the diabetic and treated PPI subnetwork

Table 5

Genes part of the proteasome complex significant expressed in the diabetic and treated group

Entrez GeneID	Symbol
5711	PSMD5
5688	PSMA7
5695	PSMB7
5690	PSMB2
10213	PSMD14
5716	PSMD10



(cont...)

(...cont)

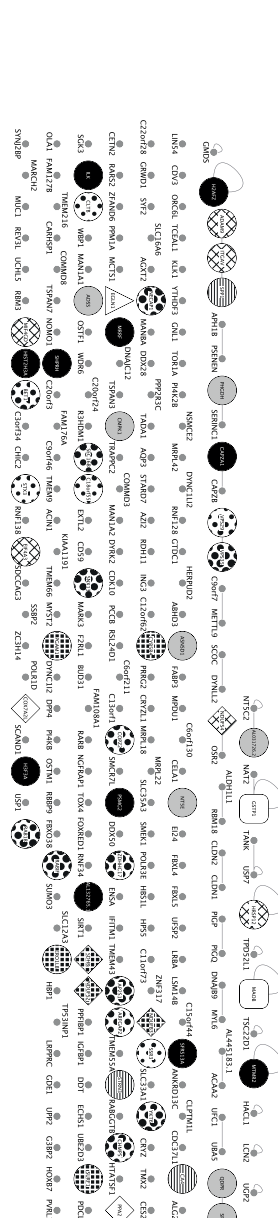
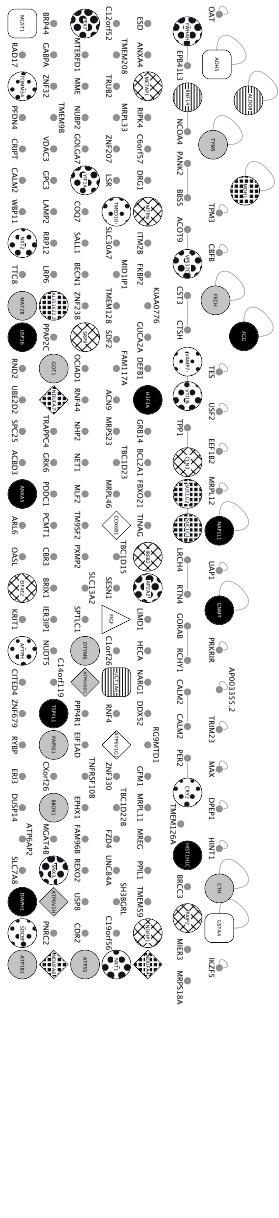
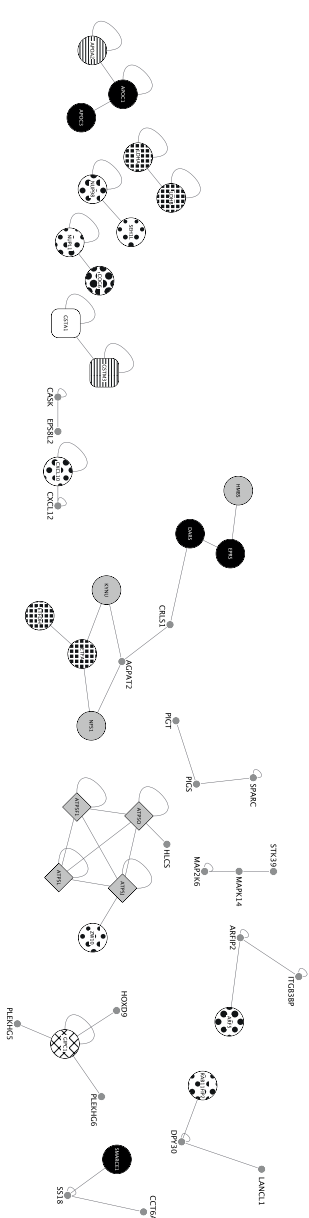


Figure 19  
Diabetic and Treated PPI subnetwork

## 5.6 Peroxisome proliferator-activated receptor- $\gamma$ coactivator-1 $\alpha$ (PGC-1 $\alpha$ )

Some biological process are not regulated by the activity of transcriptions factors instead they are control by transcriptional coactivators. Transcriptional coactivators are proteins or protein complexes that mediate the interaction between transcription factors and the transcriptional machinery. Their role in gene expression is to increase the rate of transcription by interacting with transcription factors but they do not bind directly to DNA in a sequence-specific manner (Puigserver 2003).

During the study of the regulation of the adaptive thermogenesis process, the role of nuclear receptors, such as, thyroid hormone receptors (TR) and peroxisome proliferator-activated receptors (PPAR) was elucidated. These receptors have important functions in the differentiation of brown fat cells and in the expression of the mitochondrial uncoupling proteins (UCP). The research studies also suggested that nuclear receptors cannot determine the adipocyte cell fate during adaptive thermogenesis by themselves (Handschin & Spiegelman 2006). As a consequence, it was proposed that the function of nuclear receptors was regulated by cofactors. Within this context Puigserver *et al.* (1998) cloned and characterized a novel coactivator of PPAR $\gamma$  denominated as PGC-1 (Puigserver et al. 1998). The continuous analysis of the functions of PGC-1 revealed the existence of a family of coactivators: PGC-1 $\alpha$ , previously annotated as PGC-1, (Puigserver et al. 1998), PGC-1 $\beta$  (Lin et al. 2002) and PGC-1-related coactivator (PRC) (Andersson & Scarpulla 2001).

Peroxisome proliferator-activated receptor- $\gamma$  coactivator-1 $\alpha$  (PPARGC1A or PGC-1 $\alpha$ ) is a transcriptional coactivator that regulates key biological programs that respond to the energy demands of the cell due to physical or chemical changes of the environment. These responses include: increase of mitochondrial biogenesis, cellular respiration rates, and energy substrate uptake and utilization (Finck & Kelly 2006). PGC-1 $\alpha$  is highly expressed in tissues with elevated oxidative capacity where it regulates the mitochondrial function and the cellular energy metabolism. Its expression is induced by physiological conditions that required an increased in mitochondrial energy production, such as, exercise and fasting. The activity and specificity of PGC-1 $\alpha$  is regulated at multiple different levels. The exact mechanisms that trigger the expression and activity of PGC-1 $\alpha$  are being study. The recent findings suggest that the regulatory programs that act over PGC-1 $\alpha$  include concentration of endogenous ligands, heterodimeric interaction with other coactivators, posttranslational modifications, splice variants, and autoregulatory loops (Handschin & Spiegelman 2006).

Many human pathologies, such as, neurodegenerative and heart diseases have been linked with irregular mitochondrial function. During recent years Type 2 diabetes has also been associated with mitochondrial dysfunction (Lowell & Shulman 2005). As a consequence key regulatory genes involved in the function of mitochondria have been proposed as possible target genes for the treatment of these disorders. PGC-1 $\alpha$  due to its essential activation of transcriptional programs relevant for energy metabolic pathways has been the center of study of many research groups. The roles of PGC-1 $\alpha$  in energy homeostasis, glucose metabolism, and mitochondrial activity postulate PGC-1 $\alpha$  as a target gene for the clinical treatment of diabetes.

Moreover, recent studies have associated different polymorphism in PGC-1 $\alpha$  with the risk of type 2 diabetes mellitus (Y. Yang et al. 2011). However, the complexity of the disease and the lack of a clear relationship between PGC-1 $\alpha$  and DM make it difficult to state a conclusion. The regulatory role of PGC-1 $\alpha$  in energy homeostasis and its participation on the pathophysiology of human diseases is not clear. The great regulatory capacity of PGC-1 $\alpha$  and the participation of this protein in fundamental biological responses rise new ambitions in the development of new therapeutic treatments.

The expression of PGC-1 $\alpha$  (Figure 20) fluctuates across the experimental groups. PGC-1 $\alpha$  is down regulated in the diabetic condition. A reduction by 2-fold change can be appreciated in the expression level. After drug treatment the expression of PGC-1 $\alpha$  tends to its normal level.

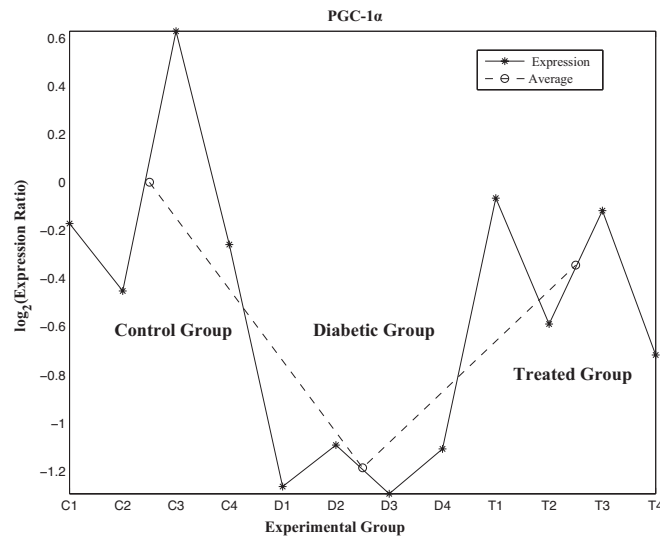


Figure 20  
Expression profile of PGC-1 $\alpha$

PGC-1 $\alpha$  as part of the complete human PPI network interacts directly with 27 proteins (Table 6). Eight of the PGC-1 $\alpha$  first neighbors are statistical annotated as

significant genes: 4 genes in the control group (CCNT1, NR1H4, ESR1, and MED24), 2 genes in the diabetic group (SFRS6 and PPARA), and 2 genes in the diabetic and treated group (SFRS5 and USF2). Moreover the genes ESRRG, ESR1 and NRF1 are annotated as transcriptional factors involved in mitochondrial biogenesis (Liang & Ward 2006). Also 7 neighbors of PGC-1 $\alpha$ : CREBBP, EP300, POLR2A, SFRS4, SFRS5, SFRS6, and CPSF2, have a degree  $k > 90$  and are consider hubs of the complete PPI network.



Table 6  
PGC-1 $\alpha$  first neighbors

Entrez GeneID	Symbol	<i>k</i>	Annotation
904	CCNT1	61	Significant gene in control group ( $q\text{-val} = 3.6 \times 10^{-04}$ )
1387	CREBBP	117	Network hub
2033	EP300	111	Network hub
2099	ESR1	87	Significant gene in control group ( $q\text{-val} = 3.6 \times 10^{-04}$ )
2104	ESRRG	11	Mitochondrial biogenesis
2908	NR3C1	78	-
3054	HCFC1	35	-
4691	NCL	73	-
4899	NRF1	15	Mitochondrial biogenesis
5430	POLR2A	202	Network hub
5465	PPARA	36	Significant gene in diabetic group ( $q\text{-val} = 4.46 \times 10^{-04}$ )
5468	PPARG	26	-
5469	MED1	45	-
6256	RXRA	64	-
6429	SFRS4	102	Network hub
6430	SFRS5	101	Network hub/Significant gene in diabetic and treated group ( $q\text{-val} = 2.11 \times 10^{-04}$ )
6431	SFRS6	98	Network hub/Significant gene in diabetic group ( $q\text{-val} < 10^{-05}$ )
7068	THRB	26	-
7392	USF2	11	Significant gene in diabetic and treated group ( $q\text{-val} < 10^{-05}$ )
7485	WRB	11	-
8648	NCOA1	59	-
9862	MED24	29	Significant gene in control group ( $q\text{-val} = 2.11 \times 10^{-04}$ )
9971	NR1H4	4	Significant gene in control group ( $q\text{-val} < 10^{-05}$ )
10060	ABCC9	15	-
23054	NCOA6	32	-
53981	CPSF2	106	Network hub
96764	TGS1	12	-

NOTE: The  $q$ -value ( $q\text{-val}$ ) represents the smallest FDR at which the gene will still consider a significant gene (Reimers 2005).

The enrichment analysis of PGC-1 $\alpha$  and its neighbors (Table 7) showed a significant enrichment in RNA biosynthetic process (GO:0032774), in transcription (GO:0006350), and in regulation of gene expression (GO:0010628) all of them with a  $p$ -value  $< 10^{-10}$ . The fold change of the three GO terms is also significant, especially for the RNA biosynthetic process with a fold enrichment F.E.  $\approx 20$ .

Table 7  
Enrichment analysis of PGC-1 $\alpha$  and its neighbors

Term	$p$ -value	F.E. *
GO:0032774~RNA biosynthetic process	$2.64 \times 10^{-12}$	20.31
GO:0006350~transcription	$5.49 \times 10^{-12}$	05.01
GO:0010628~positive regulation of gene expression	$9.22 \times 10^{-12}$	12.07

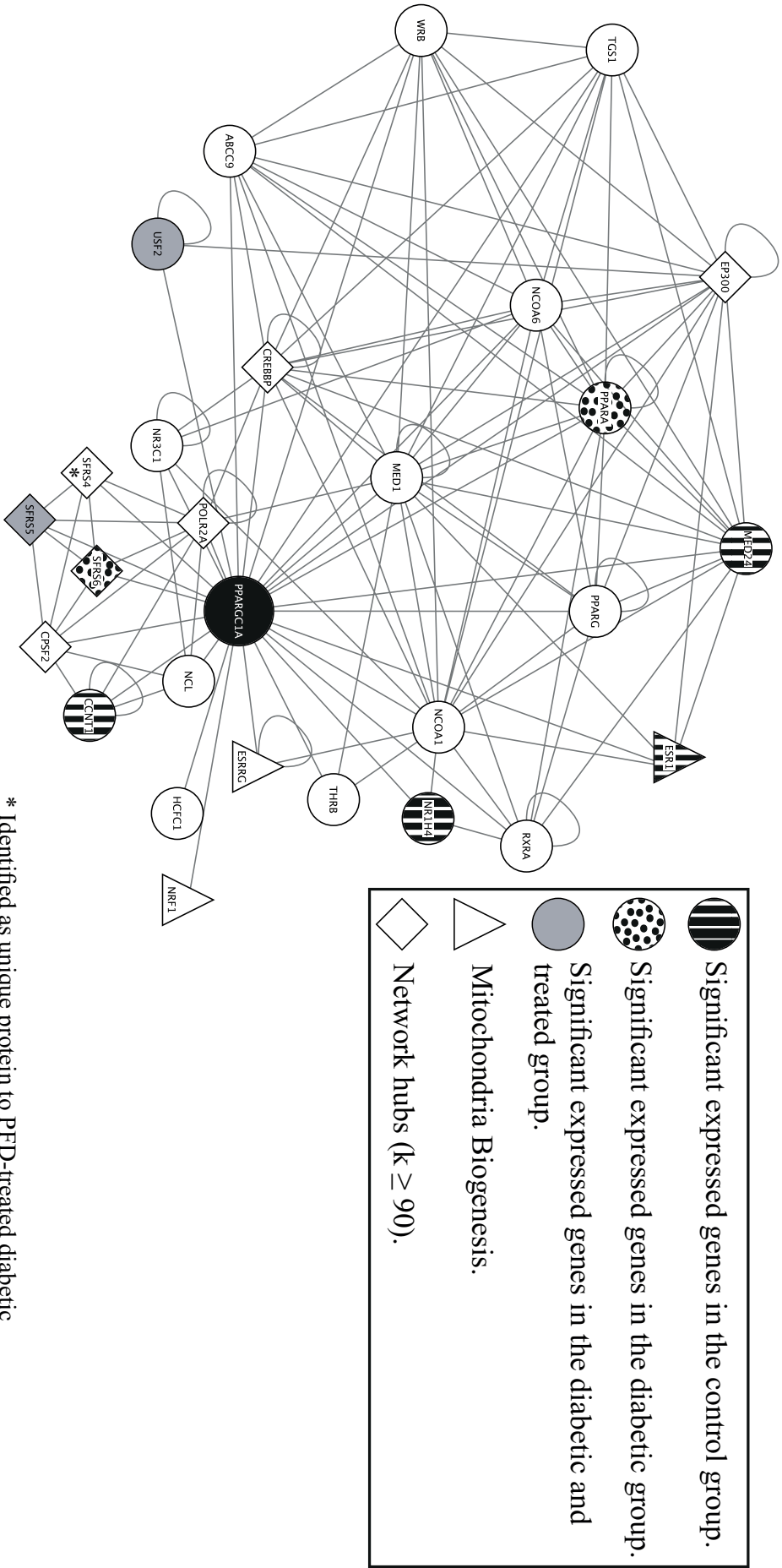
\* Fold Enrichment

The subnetwork centered on PGC-1 $\alpha$  was mapped onto the complete human PPI network. The result is a well-connected subnetwork (Figure 21) with 28 proteins and 126 interactions. The genes were highlighted according to the annotations presented in Table 6.

Out of the 27 first neighbors of PGC-1 $\alpha$  only 14 are part of the expression dataset for the present analysis. To obtain a general perspective of the expression behavior of these genes, the average of the expression ratio was plotted in  $\log_2$  scale. At the same time the expression correlation between the gene and PGC-1 $\alpha$  was visualized on the graph (Figure 22 - Figure 28).

The genes that showed a considerable dysregulation in gene expression across groups according to the expression profiles are: CCNT1 (Figure 22), ESR1 (Figure 23), and NR1H4 (Figure 28). Even though these genes have a significant correlation

with PGC-1 $\alpha$  with  $p$ -value  $< 0.05$  in all cases, only ESR1 has a Person's correlation coefficient  $|R| > 0.7$ . These 3 genes exhibited a reduction in expression in the diabetic experimental group by a 2-fold change, resembling PGC-1 $\alpha$  expression profile. However after treatment they do not have a tendency to return to their normal expression level. The genes SFRS5 and SFRS6 (Figure 26) also show great dysregulation. In contrast, the expression of these genes is promoted in the diabetic condition and they tend to normal expression levels after drug treatment. They have a significant correlation with PGC-1 $\alpha$  with  $|R| > 0.7$  and  $p$ -value  $< 0.05$ .



\* Identified as unique protein to PFD-treated diabetic kidney by RamachandraRao, S. *et al.*, 2009.

Figure 21  
PPI subnetwork centered on PGC-1 $\alpha$  (PPARGC1A)

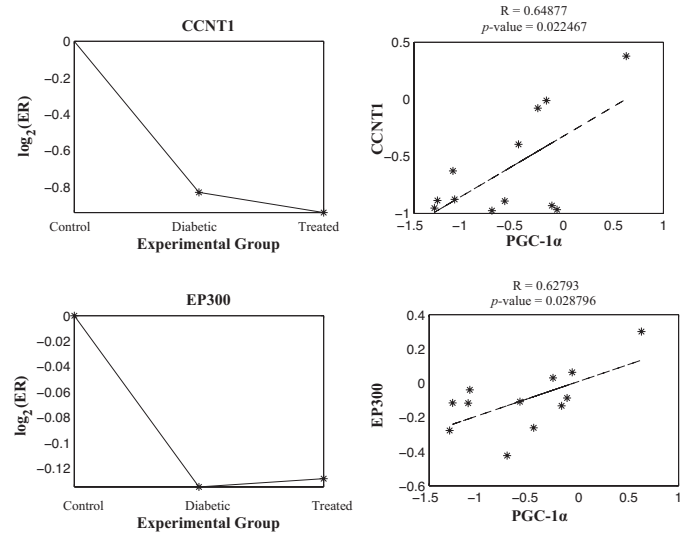


Figure 22  
Expression profile of CCNT1 and EP300 and correlation with PGC-1 $\alpha$

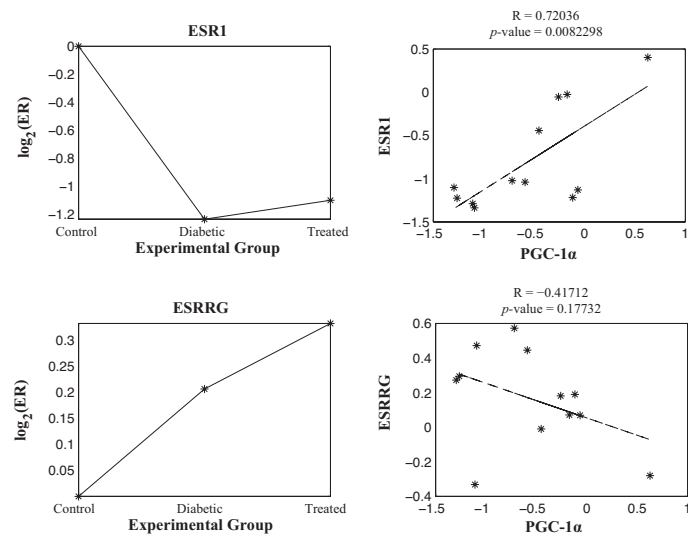


Figure 23  
Expression profile of ESR1 and ESRRG and correlation with PGC-1 $\alpha$

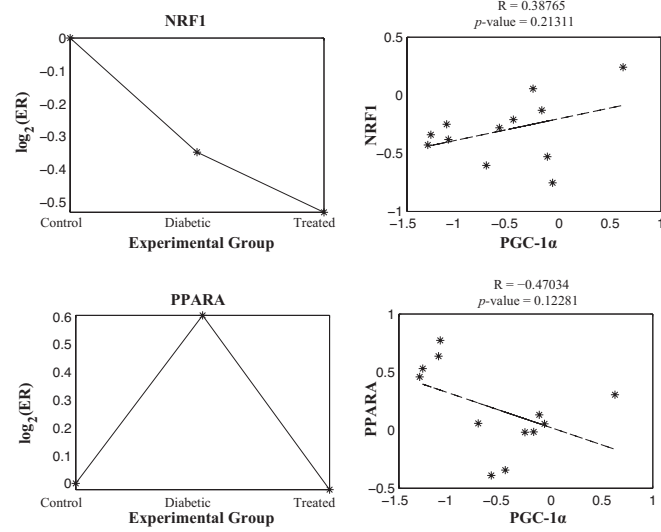


Figure 24  
Expression profile of NRF1 and PPARG and correlation with PGC-1α

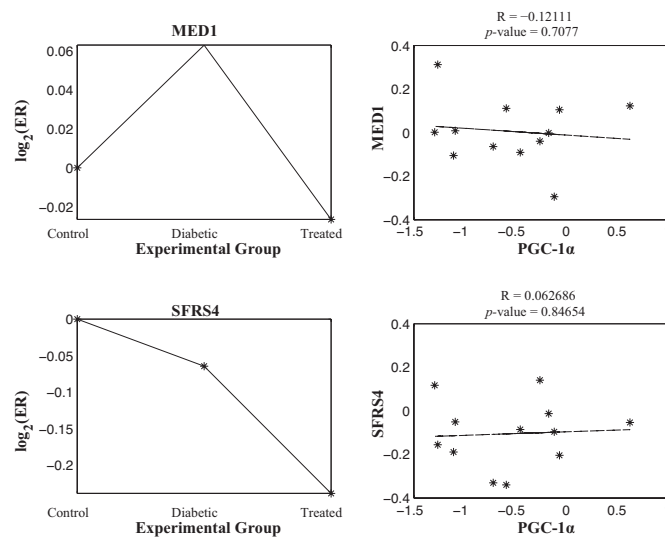


Figure 25  
Expression profile of MED1 and SFRS4 and correlation with PGC-1α

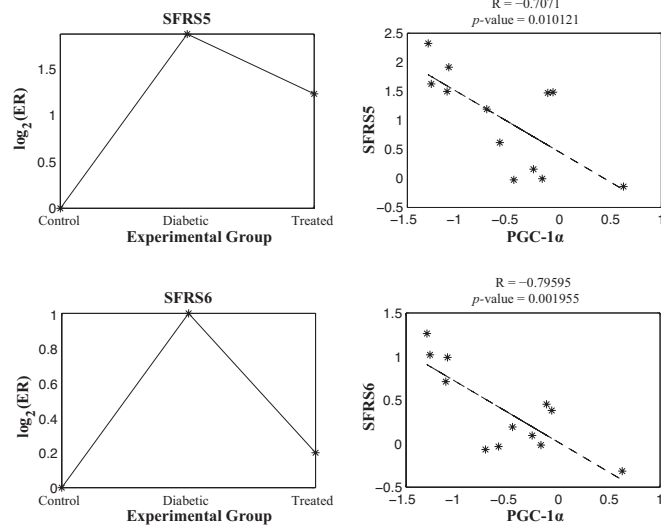


Figure 26  
Expression profile of SFRS5 and SFRS6 and correlation with PGC-1α

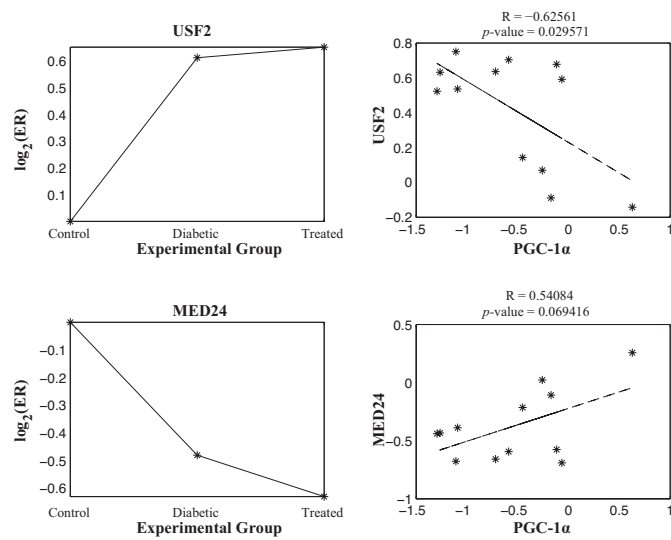


Figure 27  
Expression profile of USF2 and MED24 and correlation with PGC-1α

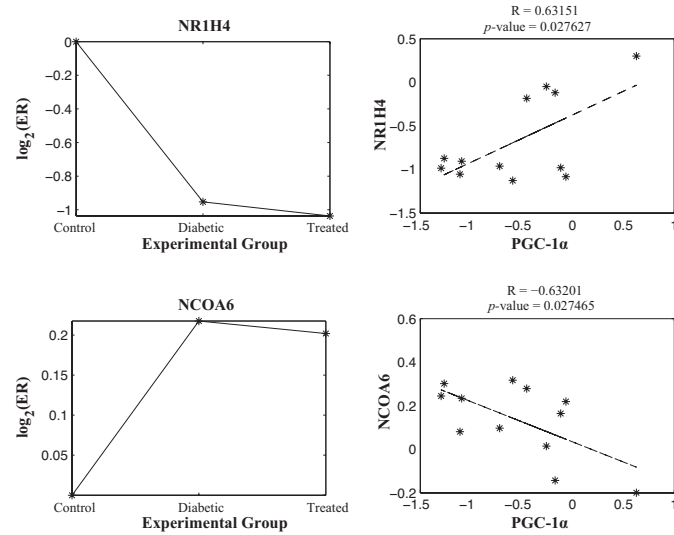


Figure 28  
Expression profile of NR1H4 and NCOA6 and correlation with PGC-1 $\alpha$

The first neighbors of PGC-1 $\alpha$  were classified in functional clusters using DAVID. Nineteen of the neighbors could be grouped in 3 different functional clusters (Table 8): Functional Cluster 1 annotated for transcriptional factors, Functional Cluster 2 annotated for coactivators, Functional Cluster 3 annotated for mRNA processing.

Table 8  
Functional clusters (FC) for PGC-1 $\alpha$  neighbors

FC 1 (F.E. = 7.27)		FC 2 (F.E. = 5.6)		FC 3 (F.E. = 3.3)	
Gene	Symbol	Gene	Symbol	Gene	Symbol
7392	USF2	8648	NCOA1	6431	SFRS6
5468	PPARG	23054	NCOA6	6430	SFRS5
7068	THRB	9862	MED24	53981	CPSF2
2104	ESRRG	5469	MED1	6429	SFRS4
5465	PPARA	8648	NCOA1	96764	TGS1
9971	NR1H4				
2099	ESR1				
6256	RXRA				
2908	NR3C1				

\* Fold Enrichment



The functional architecture of the PGC-1 $\alpha$  PPI subnetwork can be understood by analyzing the functional domains of the protein (Figure 29). PGC-1 $\alpha$  protein consists of 798 amino acids and three important functional motifs. 1) Towards the N-terminal domain of the PGC-1 $\alpha$  protein there is a LXXLL motif that is responsible for ligand-dependent interaction between coactivators and nuclear hormone receptors. This domain can also recruit proteins with histone acetyl transferase (HAT) activity but only in the presence of a transcription factor. 2) The protein also possesses a serine/arginine rich region (RS) that can interact with the C-terminal domain of RNA polymerase II. 3) At the end of the C-terminus a RNA-binding motif (RMM) is responsible for the recruit of mRNA processing proteins and splicing factor (Puigserver 2003). PGC-1 $\alpha$  is a multiple functional protein that in the presence of transcription factors it operates as a scaffold for the docking of histone modifying enzymes, the binding of the transcriptional initiation complex to the transcription factor, and the processing of the just transcribed mRNA (Handschin & Spiegelman 2006).

The functional analysis of PGC-1 $\alpha$  subnetwork showed that 33% off the neighbor's genes comprise the functional cluster 1 and are annotated as transcriptional factors; this list includes the nuclear receptors and the hormone receptors. The two remaining functional clusters, which represent coactivators and RNA processing proteins (mainly splicing factors) respectively, grouped 5 genes each. In total these 3 functional groups enclose 70% of all PGC-1 $\alpha$  firsts neighbors. Eight neighbors of PGC-1 $\alpha$  do not fit in any of the functional clusters. In this group are key proteins of the transcriptional complex: an RNA polymerase (POLR2A) and two proteins with HAT activity (CREBBP and EP300). The functional analysis of the PGC-1 $\alpha$  PPI

subnetwork elucidates the importance of PGC-1 $\alpha$  in transcription and RNA processing. The enrichment analysis complements the previous results with significant enrichment in transcription and gene expression.

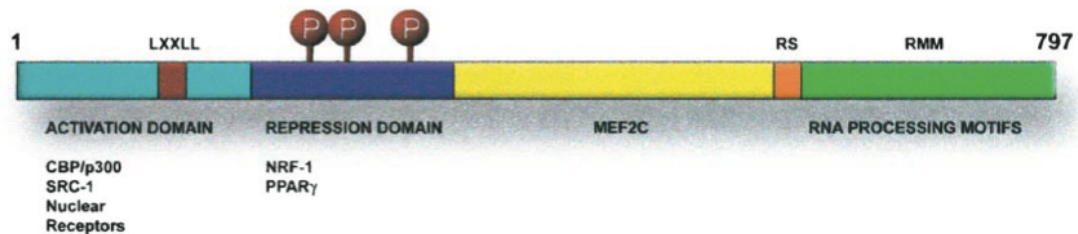


Figure 29  
Architecture of PGC-1 $\alpha$  protein  
(Puigserver 2003)

The network hubs that are part of the PGC-1 $\alpha$  subnetwork are also genes that have a direct participation in the translational process (Figure 32). Even though there is a limited expression data for these genes it is interesting to point out that the splicing factors, SFRS5 and SFRS6, are the two genes that showed the most dysregulated behavior in their expression profile (Figure 30) in comparison with the other PGC-1 $\alpha$  first neighbors. SFRS5 is significantly upregulated in the diabetic and treated group while SFRS6 is significantly upregulated only in the diabetic group. Although both genes show a tendency to return to their normal expression level after treatment only SFRS5 reached values close to the control group. This could be explained by comparing the level of expression of both genes in the diabetic mouse. SFRS6 is upregulated by a 2-fold change in comparison with the control group while SFRS5 is upregulated by a 3.5-fold change. The difference in dysregulation of expression in the diabetic experimental group could be responsible for preventing SFRS5 to return to its normal expression level.

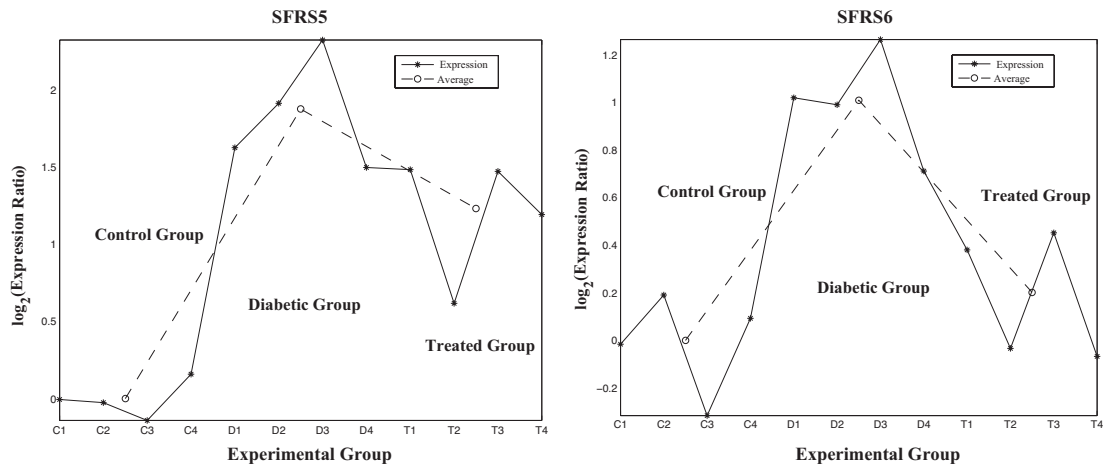


Figure 30  
Expression profile of SFRS5 and SFRS6

The remaining splicing factor, SFRS4, does not have major changes in expression (Figure 31) however this protein was previously recognized as a “PFD-unique mouse protein” by RamachandraRao *et al.* (2009). These three transcription factors together with POLR2A, PGC-1 $\alpha$ , and the polyadenylation factor (CPSF2) assemble a fully connected subnetwork, highlighted in Figure 32.

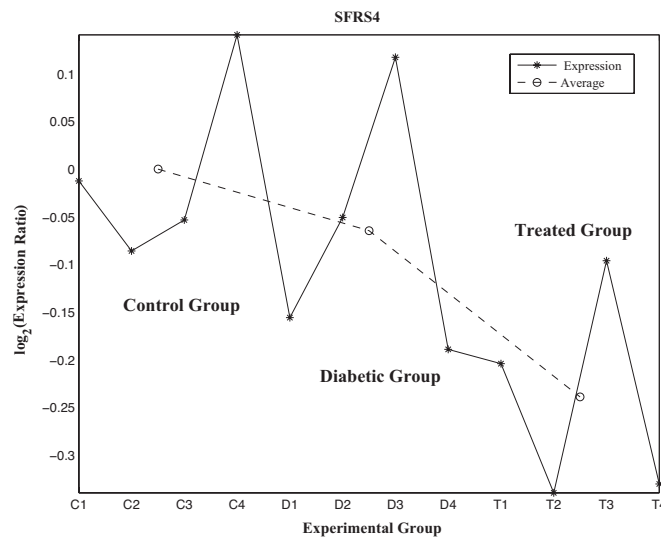


Figure 31  
Expression profile of SFRS4

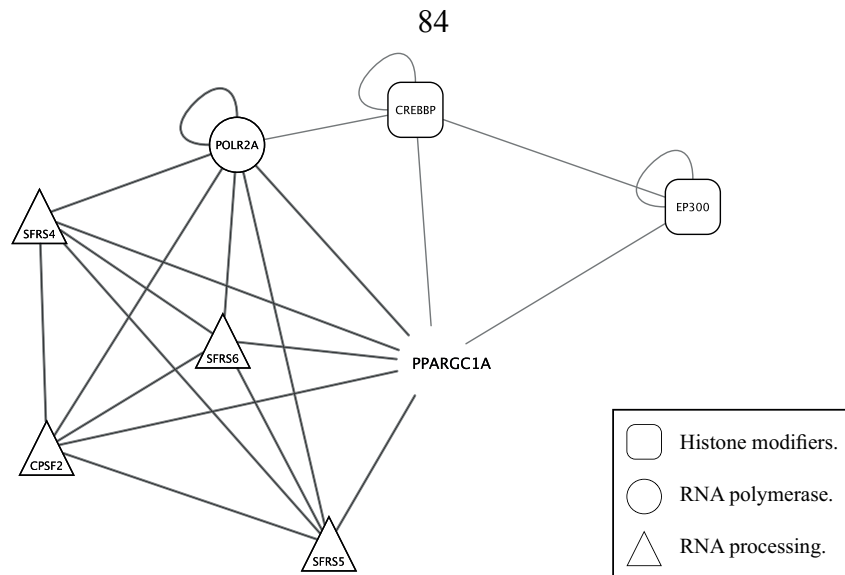


Figure 32  
Network hubs within the PGC-1 $\alpha$  subnetwork

The 4 genes involved in the production of mature mRNA annotated as RNA processing proteins share 92 genes as common neighbors (Figure 33). Approximately 80% of the proteins that interact with one of these genes also interacts with the other three. This is important because major changes in expression on one of these genes could cause a great effect in the overall function of the spliceosome machinery. This assumption is supported by the enrichment analysis of these genes and its first neighbors. SFRS4, 5, 6, and CPSF2 have a significant enrichment for RNA processing and splicing (Table 9) with an important fold enrichment that in most cases is  $> 20$ .

Figure 33

Table 9  
Enrichment analysis of SFRS4, SFRS5, SFRS5, and CPSF2

Protein	<i>k</i>	Term	<i>p</i> -value	F. E. *
SFRS5	101	GO:0008380~RNA splicing	2.04x10 <sup>-30</sup>	28.83
		GO:0006397~mRNA processing	8.51x10 <sup>-30</sup>	27.13
SFRS4	102	GO:0008380~RNA splicing	2.04x10 <sup>-30</sup>	28.83
		GO:0006397~mRNA processing	8.51x10 <sup>-30</sup>	27.13
CPSF2	106	GO:0000398~nuclear mRNA splicing (spliceosome)	3.92x10 <sup>-38</sup>	51.55
		GO:0006397~mRNA processing	6.74x10 <sup>-32</sup>	26.15
		GO:0016071~mRNA metabolic process	3.26x10 <sup>-30</sup>	22.51
SFRS6	99	GO:0008380~RNA splicing	5.14x10 <sup>-35</sup>	30.20
		KEGG:03040~Spliceosome	4.66x10 <sup>-20</sup>	18.23

\* Fold Enrichment

In total there are 21 genes annotated in the GO terms: RNA splicing and mRNA processing, in accordance to the enrichment analysis of these 4 genes (SFRS5, SFRS4, SFRS6, and CPSF2) and its first neighbors. Eighteen of these 21 genes interact with all the 4 genes (Figure 34). Ten of these genes are members of the

diabetic and treated expression cluster, 2 genes are part of the treated expression cluster, and 1 gene of the diabetic expression cluster. Meaning that more than 50% of the genes enriched in RNA processing and splicing that interact with the previously mentioned network hubs presents expression dysregulation across the experimental groups.

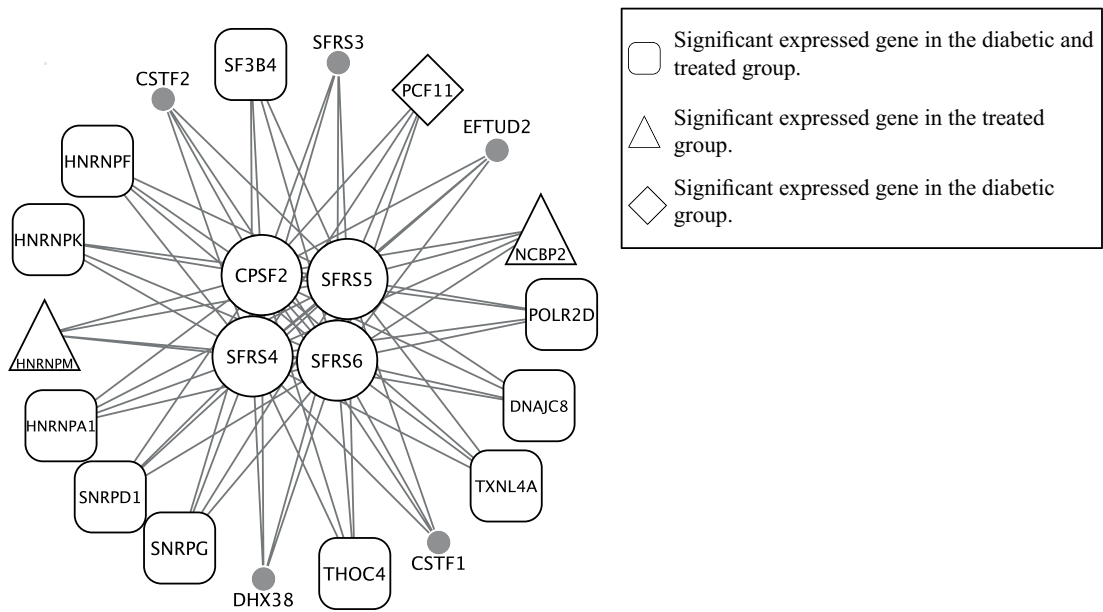


Figure 34  
PPI subnetwork of RNA processing genes

Eight of the genes presented in Figure 34 have a fold change in expression  $> 2$  in the diabetic experimental group (Figure 35). These genes have a similar expression behavior. All the genes are overexpressed in the presence of the disease and in most case, with the exception of SNRPG, when the treatment is implemented their expression tends to go down however it does not reach the normal expression value as in the control group. This behavior resembles the expression profile of SFRS5 and SFRS6 (Figure 30). When mapped onto the PPI network these genes build a fully connected network (Figure 36) meaning that all genes interact between each other.

The genes that have a function in the splicing and processing of mRNAs that present a dysregulated expression profile have a similar pattern in expression across the experimental groups. This suggests that the splicing machinery is being affected in the diabetic mouse and that the effect is not random since the expression of the genes behave in a similar fashion. It is also important to notice that the impact is done in tightly connected areas of the PPI network. The inclination of the expression of these genes towards levels similar as in the control group after drug treatment might be a consequence of the influence of PFD.

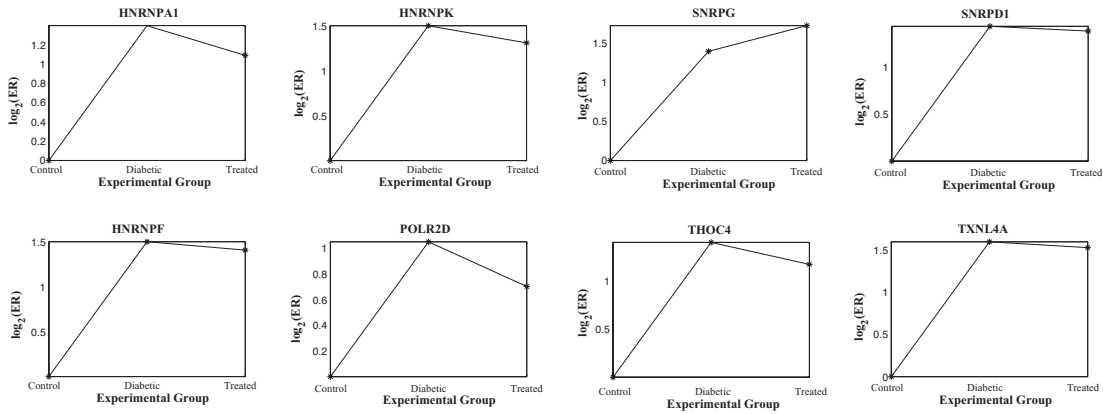


Figure 35  
Expression profile of genes involved in RNA processing

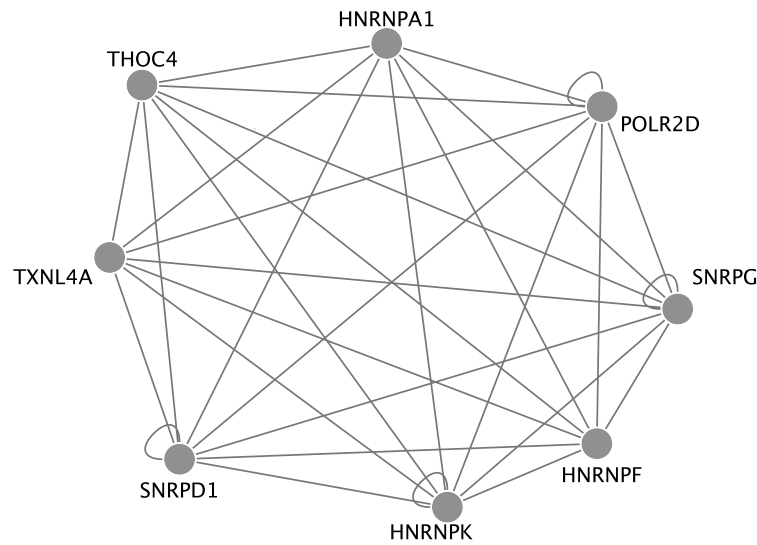


Figure 36  
PPI subnetwork centered in dysregulated genes involved in RNA processing

As mentioned before, PGC-1 $\alpha$  was discovered as a coactivator of the peroxisome proliferator-activated receptors (PPARs). Within the PGC-1 $\alpha$  subnetwork there are two members of the PPARs family: PPAR $\alpha$  and PPAR $\gamma$ . The central biological function of these receptors is the transcriptional regulation of proteins required for energy homeostasis (Yessoufou & Wahli 2010).

PPAR $\alpha$  predominate role is the promotion of fatty acid oxidation. It is highly expressed in the liver, brown adipose tissue, heart, skeletal muscle, and kidney. The expression of PPAR $\alpha$  leads to transcription of gens involved in fatty acid uptake and  $\beta$ -oxidation, resulting in a reduction in the concentration of free fatty acids (FFA) (Fruchart 2009). On the other hand, PPAR $\gamma$  promotes the storage of lipids in adipose tissue and is a crucial gene in adipocyte cell differentiation (Yessoufou & Wahli 2010). It directly controls genes with key functions of adipocytes, like lipid transport, lipid metabolism and insulin signaling (Koppen & Kalkhoven 2010).

The expression profile of PPAR $\alpha$  (Figure 37) does not show major variations across the experimental groups. It presents a 0.5-fold change overexpression in the diabetic group. Even though this variation is not consider significant it can be explain by an increase concentration of FFA, it is known that fatty acids function as endogenous ligands of PPAR $\alpha$  (Yessoufou & Wahli 2010). There is no record within the microarray data for the expression of PPAR $\gamma$ .



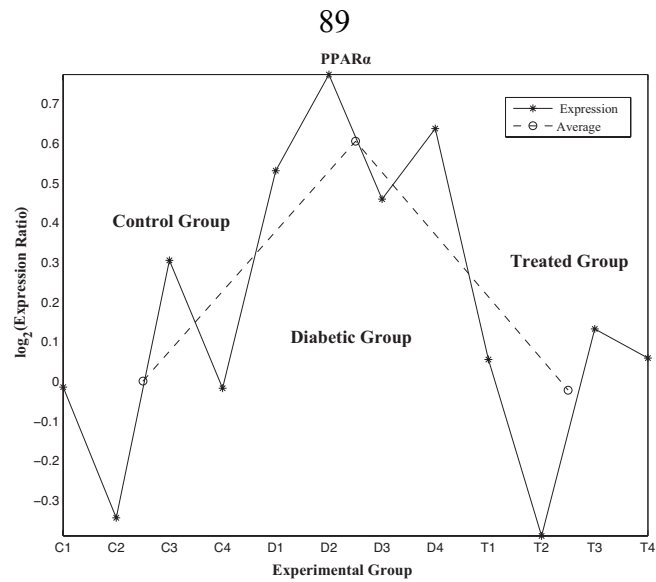


Figure 37  
Expression profile of PPARα

The two PPARs present in the PGC-1α subnetwork interact between them. They have 15 neighbors in common, including PGC-1α. Moreover, 7 of the common neighbors are also first neighbors of PGC-1α (Figure 38). The PPI subnetwork built around these two receptors has 160 interactions among 43 proteins.

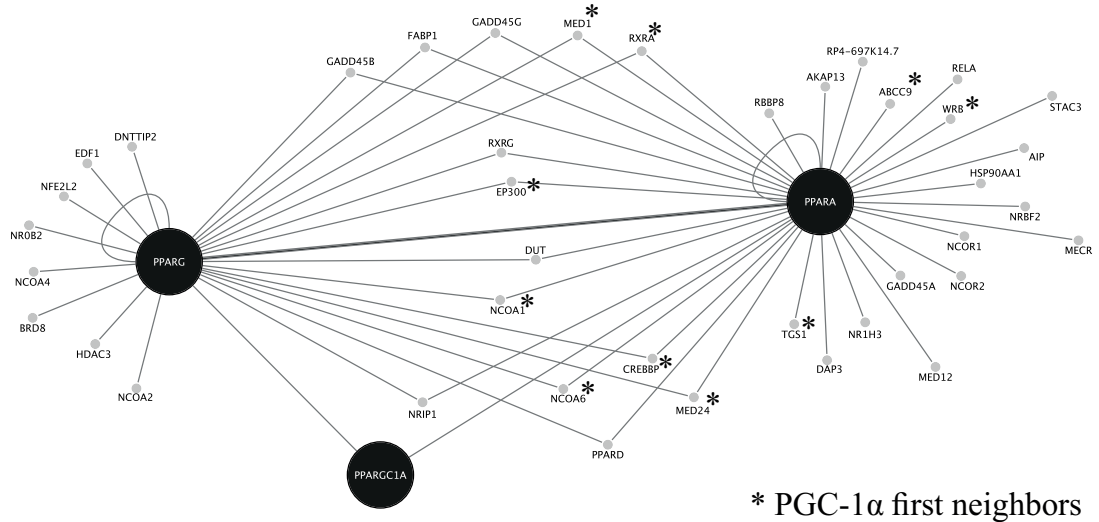


Figure 38  
PPI subnetwork centered in PPARα and PPARγ<sup>12</sup>

<sup>12</sup> Only the PPI between PPARα and PPARγ with their first neighbors are display in the figure. The interactions among the first neighbors were not visualized.

The enrichment analysis of the genes that assemble this subnetwork shows a significant enrichment in transcription and RNA metabolic process. This is expected since the majority of the genes in this network are transcription factors or proteins involved in the transcriptional machinery. There is also enrichment in signaling pathways initiated by ligand binding to a receptor, in specific, signaling pathways initiated as a consequence of steroid hormone binding (GO:0030518) and the PPAR signaling pathway (KEGG:03320) activated by fatty acids and their derivatives. The enrichment in fatty acid metabolic process and adipocytokine signaling pathway is of particular interest since high levels of FFA is a common characteristic of type 2 diabetes patients and the disturbance in fatty acid utilization and oxidation may be a determinant in the development of insulin resistance and the consequent onset of diabetes (Mensink et al. 2001).

Table 10

Enrichment analysis of PPAR $\alpha$ , PPAR $\gamma$  and their first neighbors

<b>Term</b>	<b><i>p</i>-value</b>	<b>F.E. *</b>
GO:0045449~regulation of transcription	1.39X10 <sup>-12</sup>	3.67
GO:0051252~regulation of RNA metabolic process	1.98X10 <sup>-12</sup>	4.56
GO:0030522~intracellular receptor-mediated signaling pathway	1.48X10 <sup>-11</sup>	31.18
GO:0030518~steroid hormone receptor signaling pathway	1.06X10 <sup>-10</sup>	34.58
KEGG:03320~PPAR signaling pathway	1.18X10 <sup>-07</sup>	25.57
GO:0045923~positive regulation of fatty acid metabolic process	1.68X10 <sup>-06</sup>	53.79
KEGG:04920~adipocytokine signaling pathway	1.20X10 <sup>-04</sup>	17.63

\* Fold Enrichment

The adipose tissue, in particular the white adipose tissue (WAT) besides being an energy store it actively regulates pathways responsible for energy balance. WAT secretes proteins called adipocytokines, such as, leptin, resistin, adiponectin, TNF- $\alpha$ ,

TGF- $\beta$ , and adiponectin, which are signaling molecules that activate biological programs required for the maintaining of energy homeostasis (Jazet et al. 2003).

Adiponectin an adipocyte-specific secretory protein and exclusively expressed in WAT seems to be involved in the regulation of energy balance and insulin action. It appears to be a relation between adiponectin levels in plasma and insulin resistance. Low plasma levels of adiponectin decrease insulin sensitivity and increase the risk of type 2 diabetes (Rasouli & Kern 2008). The mechanism by which adiponectin increases insulin action might be the increase of fatty acid oxidation rates causing a lower concentration of FFA and a direct improvement of insulin signaling (Snijder et al. 2006).

As shown in the diagram of the Adipocytokine Signaling Pathway (Figure 39) the presence of adiponectin (ADIPO) in the plasma and its recognition by adiponectin receptors (ADIPORs) in peripheral tissue increases the expression of PPAR $\alpha$ . The disruption of the ADIPORs leads to increased tissue triglyceride content, inflammation and oxidative stress (Shen et al. 2008). The expression of PPAR $\alpha$  promotes glucose uptake, fatty acid metabolism and the activation of  $\beta$ -oxidation pathway in mitochondria.

Although the relation between adipocytokines and type 2 diabetes is not clear and the results are not conclusive there are several studies that highlight the importance of these proteins as activating molecules of energy pathways that result in energy equilibrium and an overall increase of insulin sensitivity.

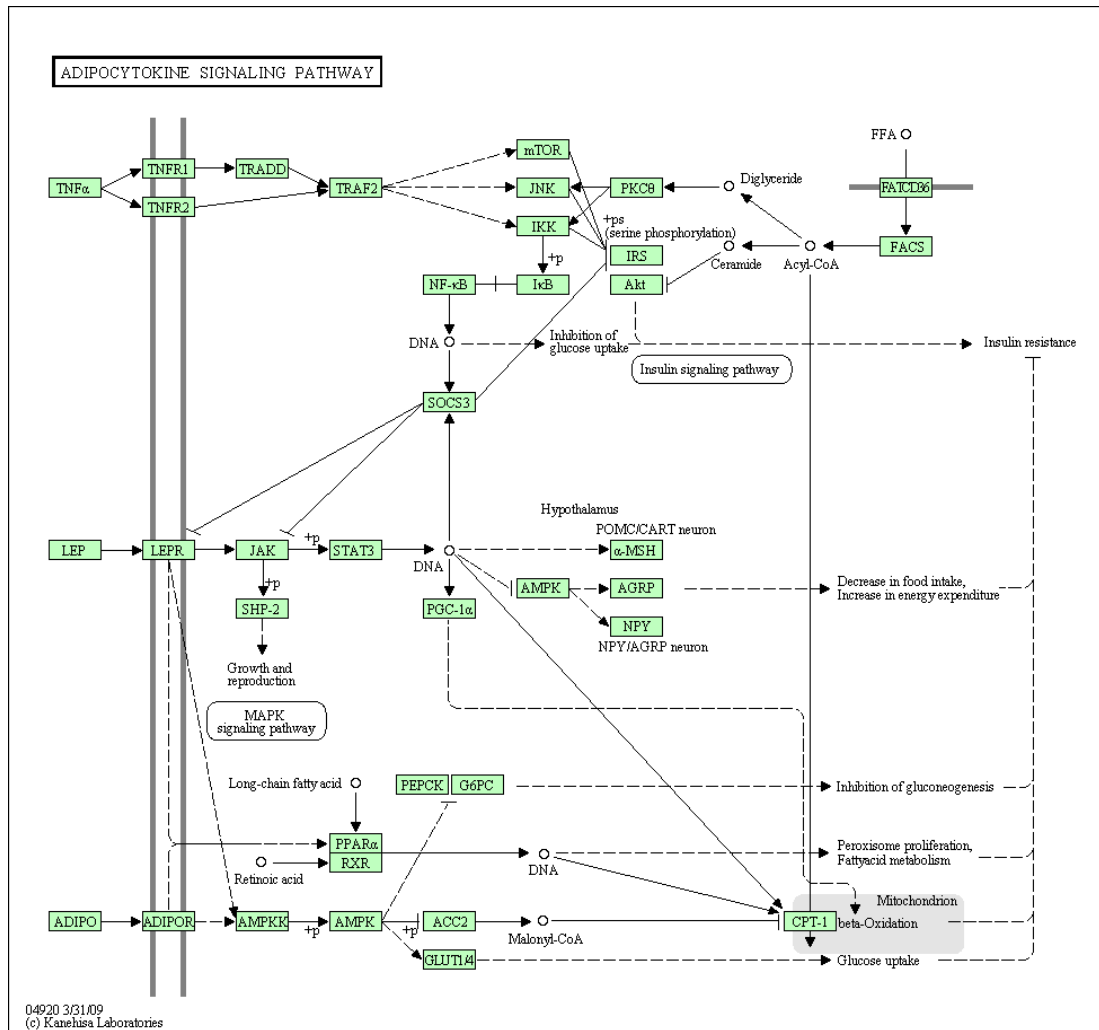


Figure 39  
Adipocytokines Signaling Pathway  
(KEGG Pathway Database)<sup>13</sup>

The PPARs in order to active transcription they required to form a heterodimer with the retinoid X receptor (RXR) (Yessoufou & Wahli 2010). The RXR-PPAR heterodimer is required to promote transcription in the presence of a ligand, either a RXR- or PPAR-specific ligand (Schulman et al. 1998). The transcriptome analysis shows a very low expression of RXR relative to the expression of PPAR $\alpha$  (Figure 40). The low expression of RXR could lead to an inactivity of PPAR $\alpha$ . Even if the gene level of PPAR $\alpha$  is normal the low concentration of RXR will result in the impediment of the formation of the functional heterodimer and as a

<sup>13</sup> <http://www.genome.jp/kegg/pathway.html>

consequence PPAR $\alpha$  will be inactive and would not be able to activate the transcription of target genes.

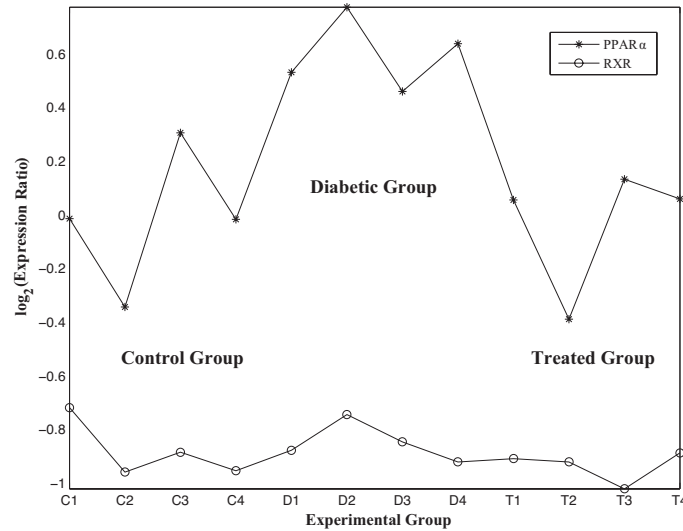


Figure 40  
Expression profile of RXR relative to PPAR $\alpha$

The estrogen receptor 1 (ESR1 or ER $\alpha$ ) is one of PGC-1 $\alpha$  first neighbors that present a major expression dysregulation. The expression profile of this gene (Figure 41) shows a decrease in expression >2-fold change in the diabetic mice. The gene keeps low expression levels after treatment and does not exhibit expression recovery. In addition to be a gene with major dysregulation across the experimental groups, ESR1 is involved in mitochondrial function and biogenesis. ESR1 as well as ESR2 are present in the nucleus and in mitochondria. The nuclear activity of ESR1 is to regulate the transcription of mitochondrial genes, specially, the nuclear respiratory factor-1 (NRF-1) in the presence of steroid hormones that act as ligands. The role of these receptors within mitochondria is less clear. The research studies suggest that ESR1 and ESR2 could be directly involve in the transcription of mitochondrial DNA (mtDNA). Another possibility is that they function as retrograde proteins involved in the communication between mitochondria and the nucleus (Klinge 2008).

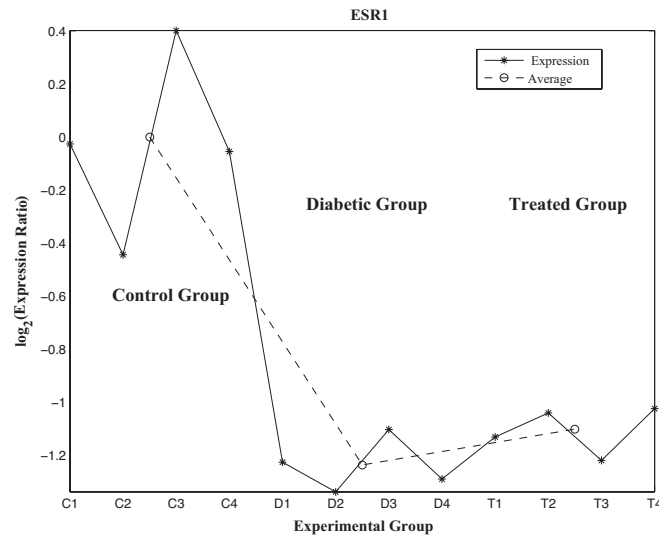


Figure 41  
Expression profile of ESR1

The estrogen-related receptors (ESRRs) are a subfamily of the steroid/nuclear receptor superfamily. They share a high protein homology with the ESRs however their activity is not ligand-dependent (Chisamore et al. 2009). There are three members of the ESRR subfamily: ESRR $\alpha$ , ESRR $\beta$  and ESRR $\gamma$ . ESRR $\alpha$  is associated with mitochondria biogenesis. Together with NRF-1 and NRF-2 it is responsible for the expression of the majority of the respiratory chain genes (Mirebeau-Prunier et al. 2010). The activity of the other two ESRRs is less evident. ESRR $\gamma$  has been associated with development and adipogenesis (Kubo et al. 2009) while the function of ESRR $\beta$  remains inconclusive.

The PPI subnetwork built around the ESRs and ESRRs (Figure 42) has 105 proteins. The ESR1 and ESRR $\gamma$  are the only estrogen receptors that interact with PGC-1 $\alpha$  however the different ESRs and ESRRs interact between each other. In some cases they interact directly, such as, ESR1 with ESRR $\alpha$ , ESR1 with ESR2 and ESRR $\alpha$  with ESRR $\beta$  but the majority of the connections between the estrogen receptors are made by an intermediate. The genes that function as links between the

different ESRs and ESRRs encode for proteins with transcriptional activity either as coactivators or as transcriptional factors. The binding of these nuclear receptors to DNA can be as homodimers or as heterodimers forming a complex with the other estrogen receptors. The formation of functional dimers requires the recruitment of basal transcription factors and coactivators that build the complete transcription machinery (Rollerova & Urbancikova 2000).

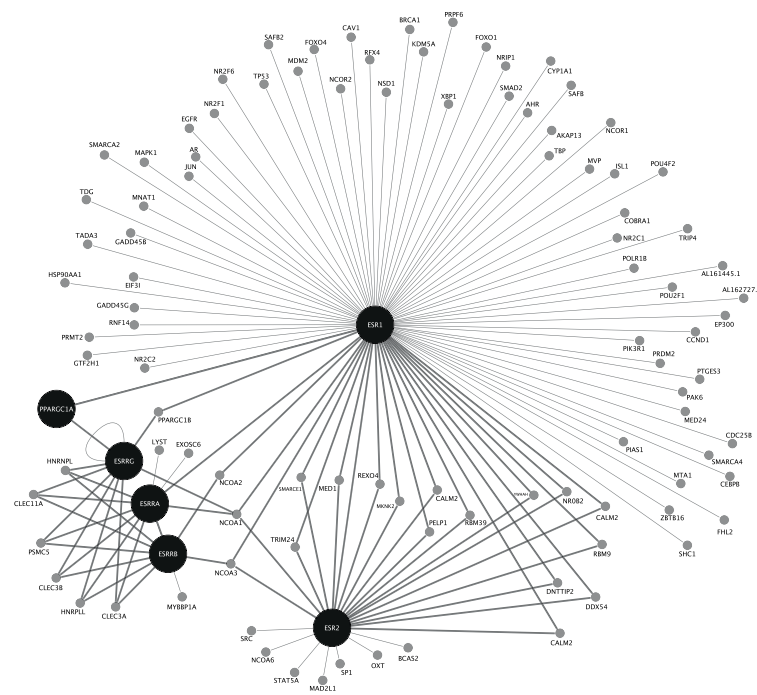


Figure 42  
PPI subnetwork centered in the ESRs and ESRRs<sup>14</sup>

The three ESRs do not show a similar expression profile (Figure 43). ESR $\beta$  and ESR $\gamma$  do not present great changes in expression across the different conditions; it can be considered that their expression is constant. However, ESR $\alpha$  expression has some variations, this gene is downregulated in the diabetic condition and the expression is recovered after treatment. The decrease in expression of

<sup>14</sup> Only the PPI between the ESRs and the ESRRs with their first neighbors are displayed in the figure. The interactions among the first neighbors were not visualized.

ESRR $\alpha$  is < 2-fold change. However the similarity with the expression profile of ESR1 makes ESRR $\alpha$  worthy of further analysis.

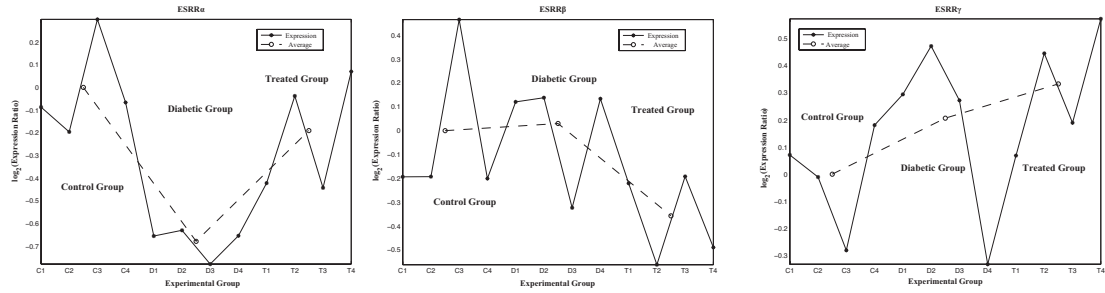


Figure 43  
Expression profile of ESRR $\alpha$ , ESRR $\beta$  and ESRR $\gamma$

The expression similarity between ESR1 and ESRR $\alpha$  is important because these two genes due to its homology, especially in the DNA binding domain, can activate the transcription of the same genes (Chisamore et al. 2009). ESRR $\alpha$  interacts with some of the coactivators that regulate the transcriptional activity of ESR1. In the absence of an exogenous ligand, ESRR $\alpha$  has the potential to regulate genes that are target for the complex ESR1-ligand. In addition, the expression ESRR $\alpha$  can be regulated by the binding of ESR1 to the multi hormone response element (MHRE) found in the promoter region of ESSR $\alpha$  (D. Liu et al. 2003). The significant expression correlation between these two genes (Figure 44) manifests the expression regulation of ESRR $\alpha$  by ESR1 and complement role that ESRR $\alpha$  plays in the transcription of target genes for ESR1 in the absence of a ligand.



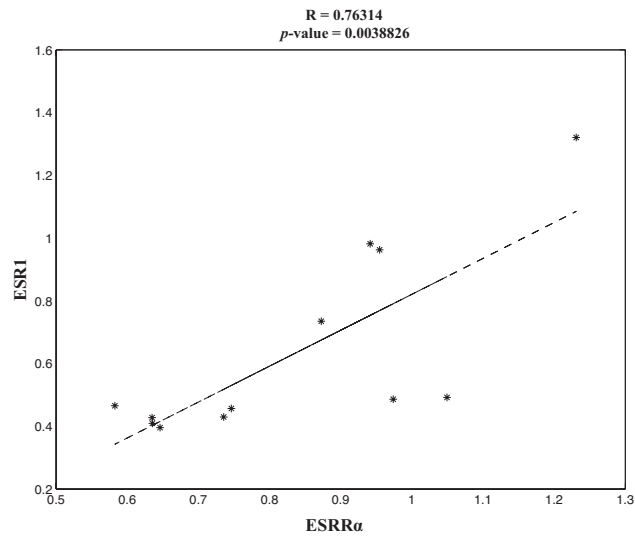


Figure 44  
Expression correlation between ESR1 and ESRRα

The downregulation of ESRRα in the diabetic mice can be a consequence of the low levels of ESR1 and its upregulation after treatment can be due to the activation of regulatory pathways by the drug that compensate for the low activity of ESR1. ESRRα can be working as a support gene to ensure that the metabolic pathways, particularly the mitochondrial oxidative function, that are regulated by ESR1 do not present major dysfunctions due to a low level of ESR1 or endogenous ligand.

The analysis of ESR1 and ESRRα is relevant since, as mentioned before, these two genes regulate the expression of two transcription factors, NRF-1 and NRF-2, which are essential for the expression of nuclear genes governing mitochondrial respiratory functions.

The expression of the respiratory systems depends on the transcription of nuclear and mitochondrial genes. The mtDNA encodes for 22 tRNAs, 2 rRNAs, and 13 subunits of respiratory complexes I, II, IV, and V. The coding limitations of

mtDNA are fulfilled by the transcription of nuclear genes. The majority of respiratory proteins, all the proteins members of the mitochondrial translation complex, and the proteins involved in the transcription and translation of the mitochondrial genome are encoded in the nucleus (Scarpulla 2008a). NRF-1 and NRF-2 are implicated in the expression regulation of nuclear genes essential for the mitochondrial respiratory apparatus (Figure 45). In addition, promoters for most of the nuclear genes encoding mtDNA transcription and replication factors have functional domains recognized by the NRFs.

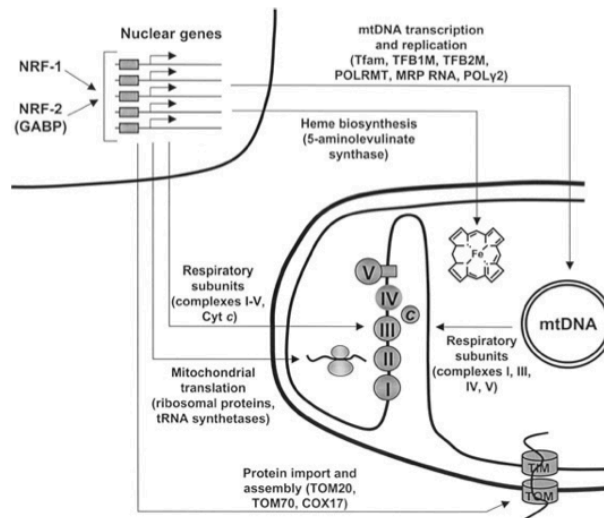


Figure 45  
Diagram of the expression of mitochondrial genes  
(Scarpulla 2008a)

The expression of NRF-1 (Figure 46) does not have major changes across the experimental groups. Nevertheless it is important to mention that the gene has a tendency to be downregulated in the diabetic group as well as in the treated group. There is no expression data for NRF-2.

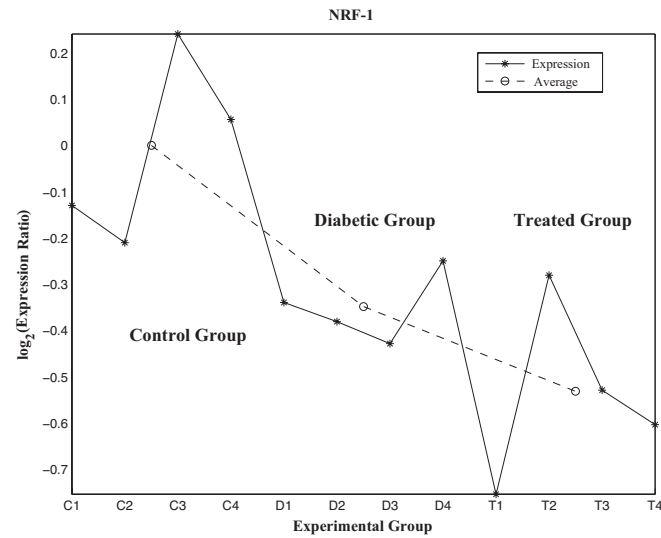


Figure 46  
Expression profile of NRF-1

The transcriptional activity of NRF-1 and NRF-2 is not a mutually exclusive event. Both factors can activate the expression of the same genes and some of these genes, including the mitochondrial transcription factors, required the presence of the two nuclear respiratory factors in order to be expressed (Scarpulla 1996). The PPI subnetwork built around these two transcription factors (Figure 47) besides being significantly enriched in transcription is annotated for the ontology groups: mitochondrion organization and homeostatic process (Table 11). The enrichment analysis shows that the interaction of NRF-1 and NRF-2 with the help of other transcription factors and histone modifying proteins regulates the expression of genes involved in the organization of mitochondria. The enrichment in the homeostatic process suggests that in addition to the regulatory role in the architecture of the mitochondria, these transcription factors regulate the mitochondrial function in order to assure the maintenance of an energy stable equilibrium.

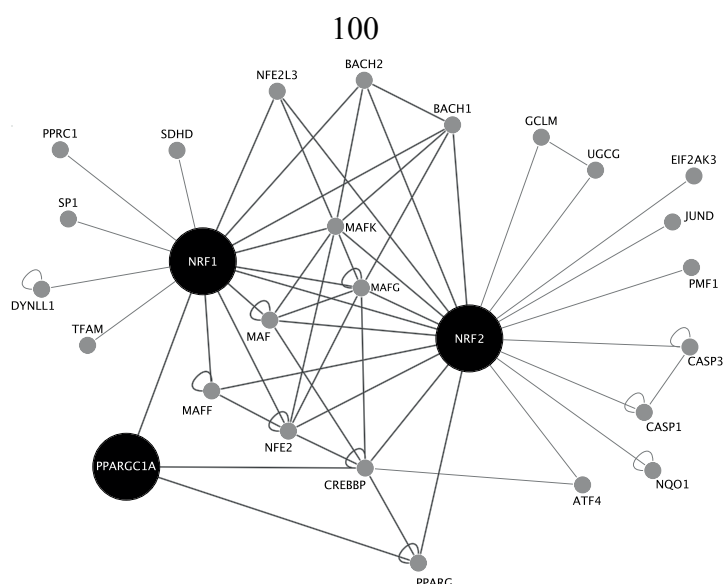


Figure 47  
PPI subnetwork centered in NRF-1 and NRF-2

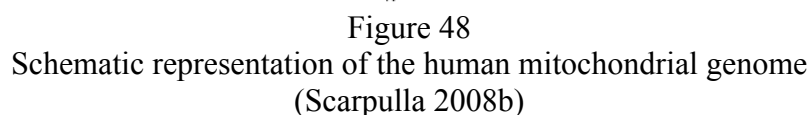
Table 11

Enrichment analysis of NRF-1, NRF-2 and their first neighbors

Term	<i>p</i> -value	F.E. *
GO:0006350~transcription	$1.15 \times 10^{-09}$	4.53
GO:0007005~mitochondrion organization	$2.27 \times 10^{-03}$	14.52
GO:0042592~homeostatic process	$2.54 \times 10^{-03}$	4.67

\* Fold Enrichment

Within this subnetwork are other two transcription factors that have been associated with mitochondrial function: the specificity protein 1 (SP1) and the transcription factor A, mitochondrial (TFAM). SP1 is a general transcription factor that regulates the expression of many different genes and has been associated with the regulation and coordination of nuclear-encoded mitochondrial genes, in specific, genes without a NRF binding domain (Ben-Shachar 2009). TFAM is encoded in the nucleus and its expression is NRFs-dependent. This transcription factor is essential for efficient and correct initiation of the mtDNA transcription. In addition to its transcriptional activity, TFAM has a central role in the mtDNA maintenance (Garstka 2003).



The expression profile of TFAM (Figure 49) presents a  $> 3$ -fold overexpression in the diabetic state and the expression does not decrease after drug treatment. TFAM is significantly expressed in the diabetic and treated group ( $q$ -value  $< 10^{-4}$ ) and has a negative correlation with PGC-1 $\alpha$  ( $R = -0.76$   $p$ -value =  $3 \times 10^{-3}$ ). The high expression of TFAM indicates that there is an increase in the mtDNA copy number in the diabetic mouse. However is not possible to conclude that the transcription of the mitochondrial genome and its respiratory capacity is also upregulated since the copy number control of TFAM is independent of its transcriptional activity (Scarpulla 2008b).

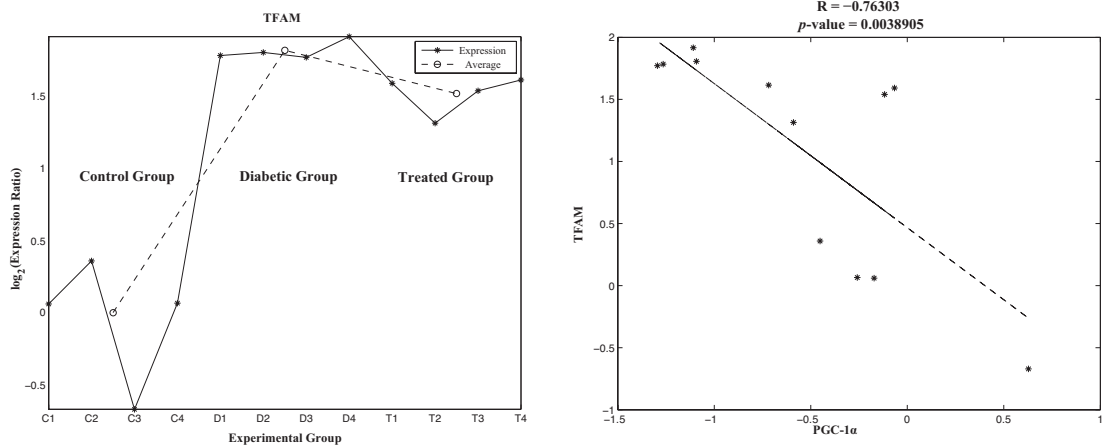


Figure 49  
Expression profile of TFAM and expression correlation with PGC-1 $\alpha$

TFAM has a low number of interactions. This transcription factor only interacts with 6 proteins (Figure 50). However, these interactions seem to be very selective. Four out of the 6 first neighbors of TFAM are genes involved in the transcriptional machinery of the mtDNA: the polymerase RNA mitochondrial (POLRMT), the transcription factor B1, mitochondrial (TFB1M), the transcription factor B2, mitochondrial (TFB2M), and the previously discussed NRF-1.

The enrichment analysis (Table 12) confirms the participation of TFAM and its first neighbors in the organization of mitochondria and the transcription of its genome. The ontology term: transcription from mitochondrial promoter (GO:0006390) is considerably enriched in this group of genes presenting a  $F.E. > 900$ . This result confirms that the subnetwork centered in TFAM plays an important role in the transcription of mtDNA and dysregulations of these genes may have a significant impact in the overall function of the organelle. Moreover, all the genes within the subnetwork have a connectivity  $k < 16$ , in specific, POLRMT, TFB1M, TFB2M, and MTF1 have a connectivity  $k = 3-4$ . The low connectivity indicates that the biological function of these genes is very specific. They only interact with a particular set of genes that usually share common biological functions and facilitate the correct operation of the gene.

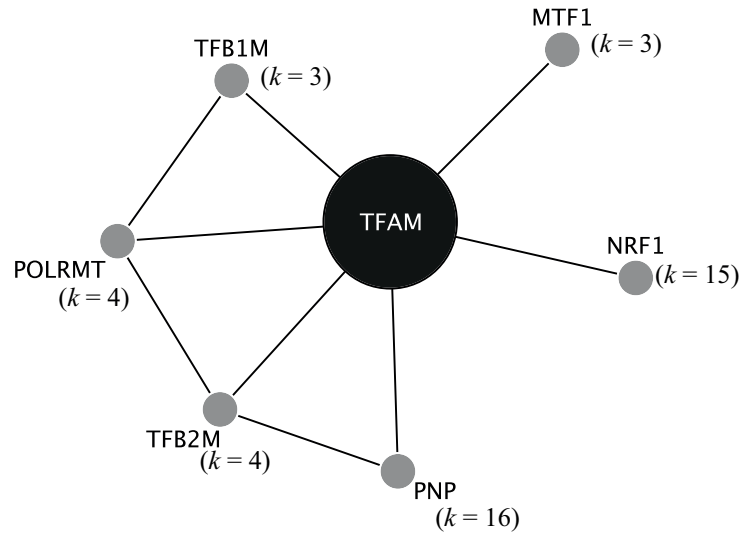


Figure 50  
PPI subnetwork centered in TFAM

Table 12  
Enrichment analysis of TFAM and its first neighbors

Term	<i>p</i> -value	F.E. *
GO:0006390~transcription from mitochondrial promoter	$2.46 \times 10^{-06}$	966.29
GO:0007005~mitochondrion organization	$2.03 \times 10^{-05}$	56.02
GO:0006350~transcription	$4.70 \times 10^{-04}$	5.52

\* Fold Enrichment

TFB1M and TFB2M are two isoforms of the mitochondrial transcription specificity factor. Both proteins together with TFAM and mitochondrial RNA polymerase (POLRMT) are essential for the proper initiation of the bidirectional transcription of mtDNA (Gleyzer et al. 2005). The promoter region of both genes has a NRFs recognition site suggesting a strong relation between the NRFs and the regulation of the two mitochondrial transcription factors. TFB1M and TFB2M have similar functions; both can transcribe the mtDNA in the presence of TFAM and POLRMT. However, TF2BM is two orders of magnitude more active than TFB1M in transcription (Asin-Cayuela & Gustafsson 2007).

The expression of TFB2M (Figure 51) remains constant across the experimental groups, even though a tendency to be downregulated can be appreciated. The changes in expression are not considerable so is not possible to state that the expression of this gene is varying. In contrast, the expression of TFB1M is downregulated in the diabetic group and returns to its normal expression levels in the treated group. Even so, it is not possible to argue that the transcription of mtDNA is being affected due to two reasons: 1) as mentioned before, TFB2M has a higher transcriptional activity and 2) RNAi knock-down experiments of TFB1M in *Drosophila melanogaster* does not result in less mitochondrial RNA transcripts but, instead, reduces mitochondrial protein synthesis (Matsushima et al. 2005).



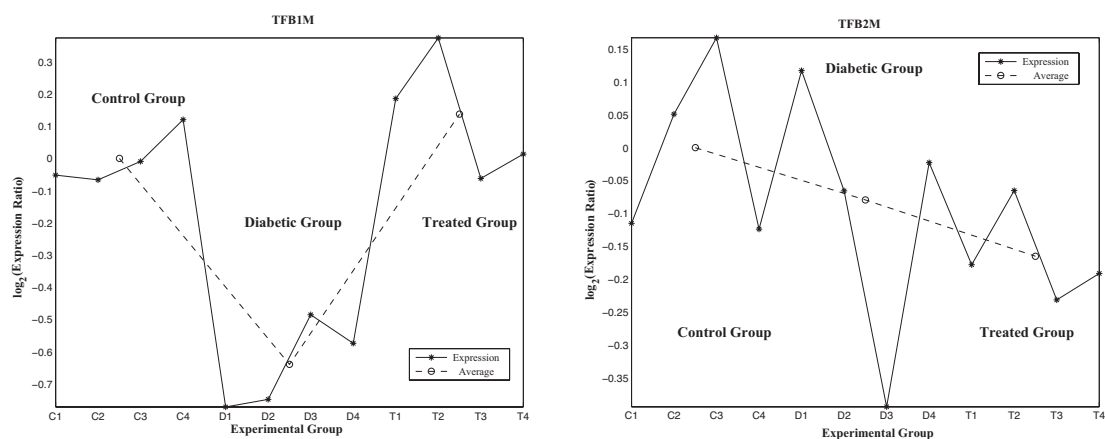


Figure 51  
Expression profile of TF1BM and TF2BM

POLRMT is a key molecule of the mitochondrial transcription machinery. This protein assembles at the promoter region to the transcriptional complex that initiates transcription of mtDNA. However, POLRMT cannot bind directly to the promoter region and as a consequence in cannot initiate transcription by itself. POLRMT depends on TFAM and one of the two mitochondrial transcription specificity factors: TF1BM or TF2BM to activate transcription (Fukuoh et al. 2009).

The expression profile of POLRMT (Figure 52) suggests that the mitochondrial translation apparatus activity might be reduced in the diabetic state due to low concentration of this protein. The expression of POLRMT is downregulated by a 2-fold change in the diabetic group and the expression is not recovered after treatment. The presence of POLRMT is a prerequisite for the transcription of mtDNA.

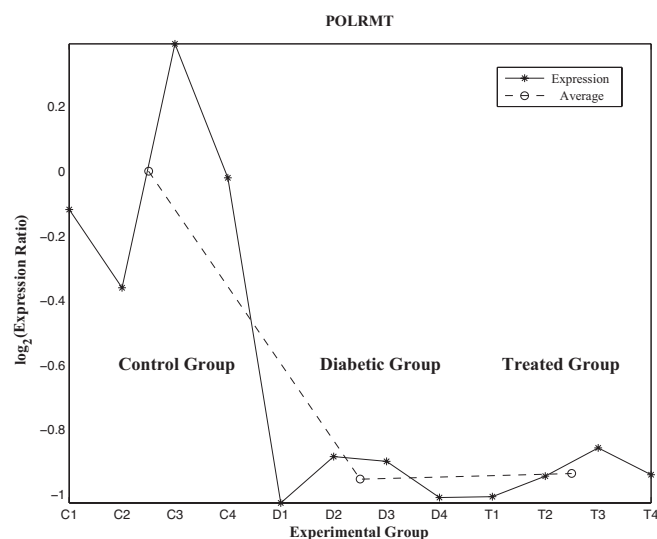


Figure 52  
Expression profile of POLRMT

The previous results show that the transcription of the mitochondrial genome could be affected due to low expression levels of important genes involved in the mitochondrial transcription machinery, such as, POLRMT and TF1BM. However, it is not clear if this has an impact in the activity of mitochondria. The genes annotated for OXPHOS and fatty acid metabolism were analyzed to observe if the expression of these genes is affected and predict a possible disruption of the mitochondrial functions.

## 5.7 Fatty Acid Metabolic Pathway and Oxidative Phosphorylation

The genes annotated for KEGG Pathways: oxidative phosphorylation (OXPHOS) (hsa00190) and fatty acid metabolism (hsa00071) were obtained using the two databases for mitochondria discussed in the Material and Methods section. This is done in order to examine how the two major biological functions of mitochondria are been affected at transcription level. The genes were first mapped onto the human PPI network. For visualization purposes only the interactions between the genes of

interest and their respective first neighbors are display. The interactions among the first neighbors are not displayed in order to have a more understandable image.

For the lipid metabolism pathway, a total of 53 genes were identified between the two databases. Only 41 genes could be mapped onto the PPI network. The subnetwork was built using the annotated genes and their first neighbors (Figure 53). The genes with significant expression in any of the experimental groups were identified. From the genes annotated in the lipid metabolism pathway there are 7 significant expressed in the control group, 9 in the diabetic and treated group and 1 in the control and treated group (Table 13).

Table 13  
Significant expressed genes annotated in lipid metabolism pathway

<b>Entrez GeneID</b>	<b>Symbol</b>
<b>Significant Expressed in the Control and Treated Group</b>	
3284	HSD3B2
<b>Significant Expressed in the Diabetic and Treated Group</b>	
2110	ETFDH
1892	ECHS1
23597	ACOT9
2108	ETFA
33	ACADL
5096	PCCB
10449	ACAA2
3032	HADHB
128	ADH5
<b>Significant Expressed in the Control Group</b>	
224	ALDH3A2
1376	CPT2
2109	ETFB
10948	STARD3
1891	ECH1
3158	HMGCS2
2639	GCDH

In addition the genes that present a dysregulation in expression  $\geq 2$ -fold change where plotted in a  $\log_2$  scale (Figure 54 and Figure 55).

For the oxidative phosphorylation pathway the databases identified 127 genes including 9 genes encoded by the mitochondrial genome (Table 14). When mapped into the PPI network, 5 subnetworks were identified that have specific biological functions (Figure 56). One hundred and three genes annotated in the oxidative phosphorylation are part of the complete PPI network; only the 5 subnetworks are display in Figure 56, the remaining genes do not interact between them.

Table 14  
Mitochondrially encoded genes annotated in the oxidative phosphorylation

Entrez GenID	Symbol
4509	MT-ATP8
4513	MT-CO2
4514	MT-CO3
4519	MT-CYB
4535	MT-ND1
4536	MT-ND2
4537	MT-ND3
4538	MT-ND4
4541	MT-ND6

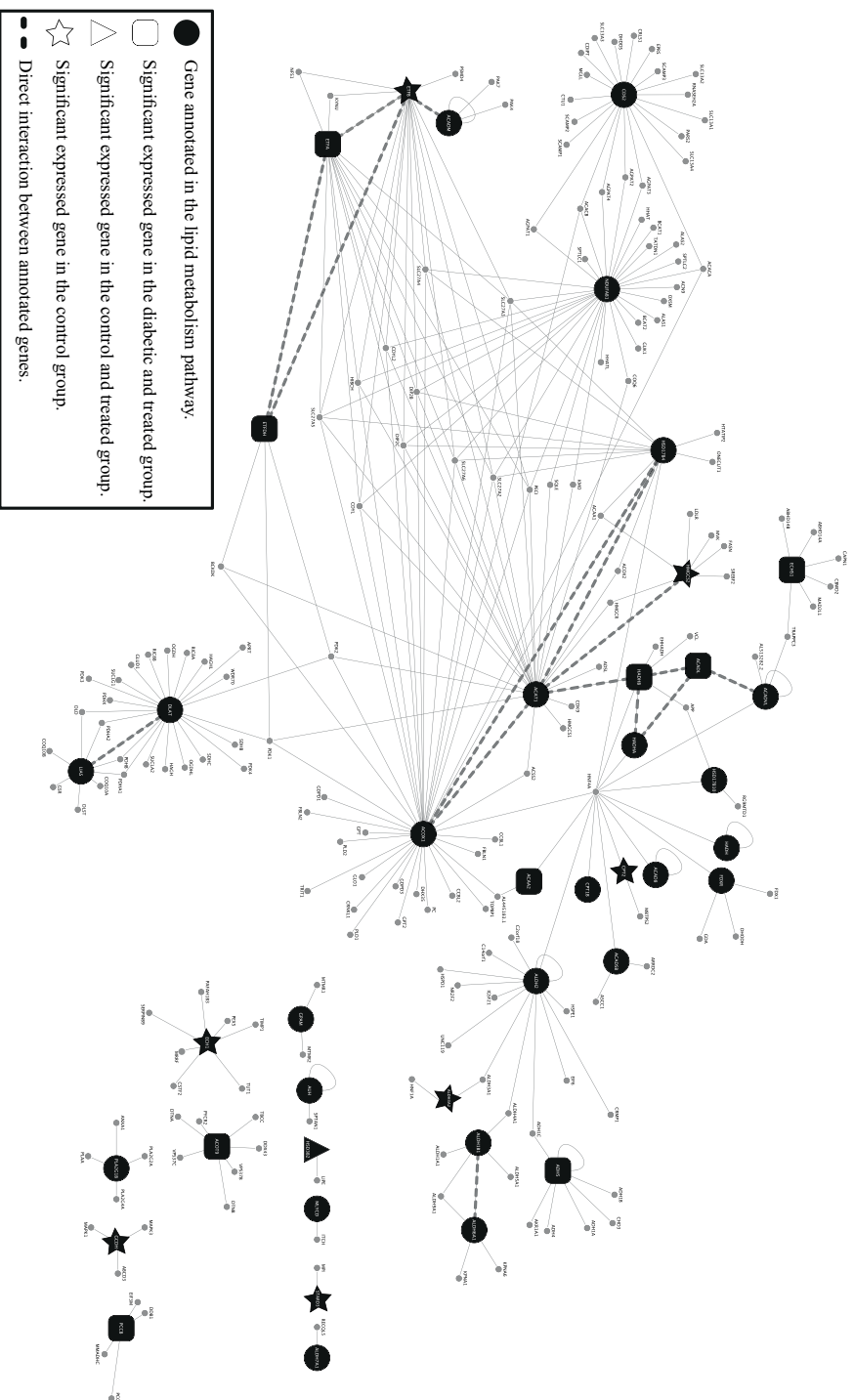


Figure 53  
 PPI subnetwork of genes annotated in the lipid metabolism pathway and their first neighbors

NOTE: Only the PPI between the annotated genes with their first neighbors are display in the figure. The interactions among the first neighbors were not visualized.

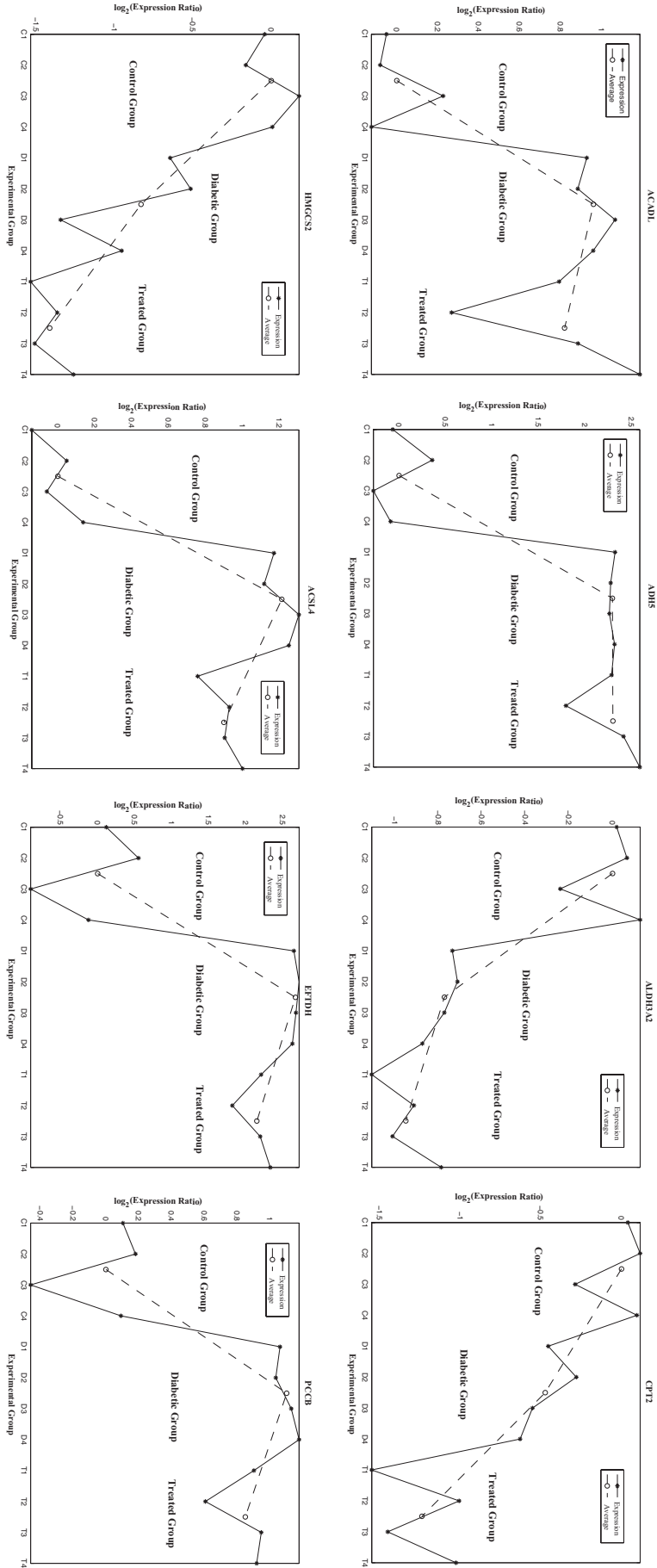


Figure 54  
Dysregulated genes annotated in lipid metabolism pathway

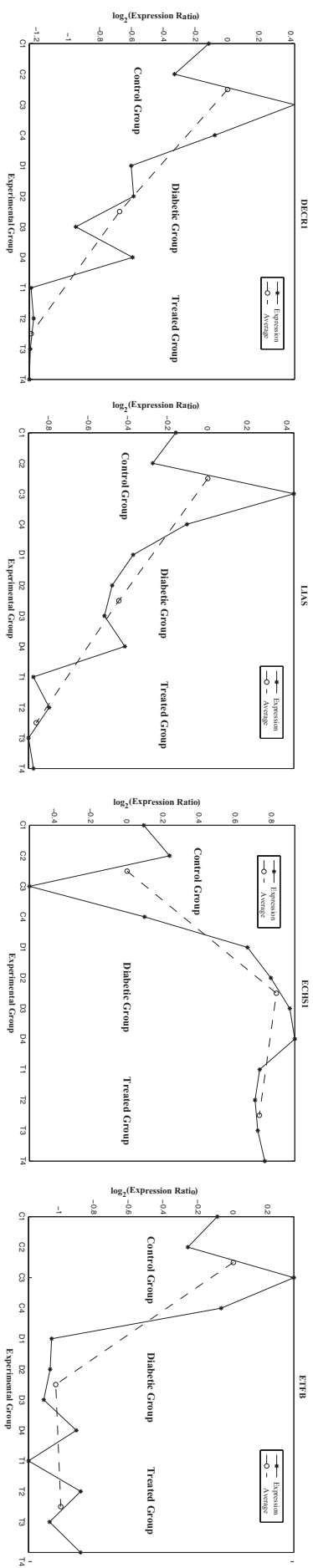


Figure 55  
Dysregulated genes annotated in lipid metabolism pathway

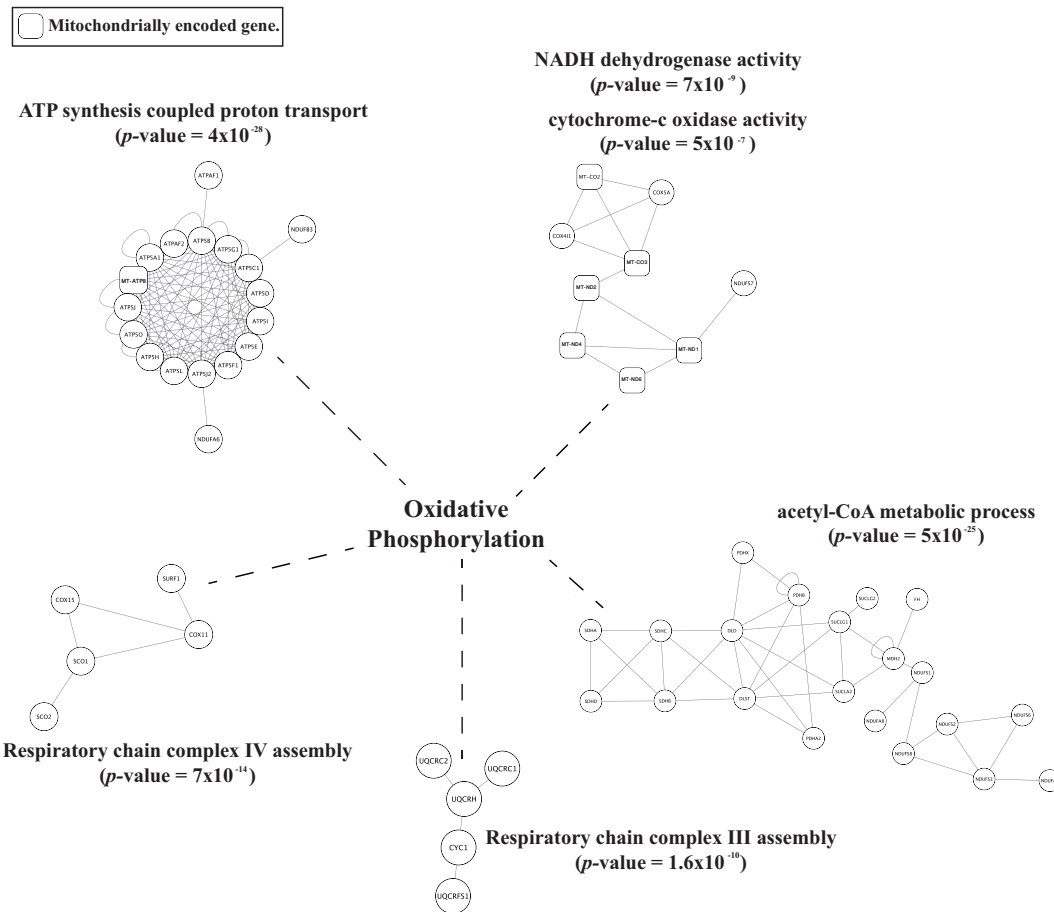


Figure 56

PPI subnetworks built with the genes annotated in oxidative phosphorylation

In relation to the statistical analysis, 24 genes are significant expressed in the diabetic and treated group and 4 in the control group. The genes significant expressed in the diabetic and treated group represent about 25% of this group of genes in particular. The expression of these genes is dysregulated in the diabetic mouse and after drug treatment do not return to normal expression levels, 16 of these genes have a dysregulation  $\geq 2$ -fold (Figure 57 and Figure 58). It is important to notice that, with the exception of BCS1L, all the dysregualted genes are overexpressed in the diabetic state.



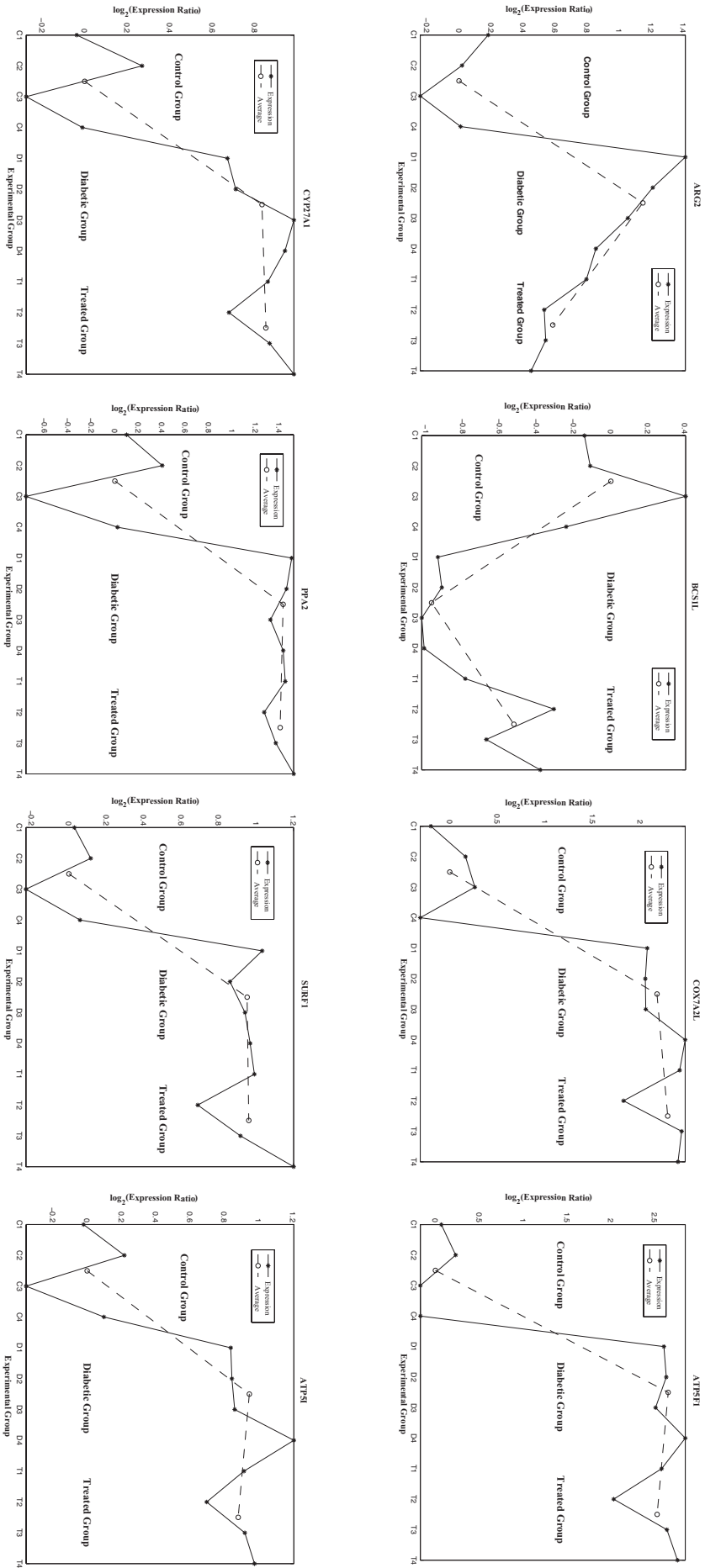


Figure 57  
Dysregulated genes annotated in oxidative phosphorylation pathway

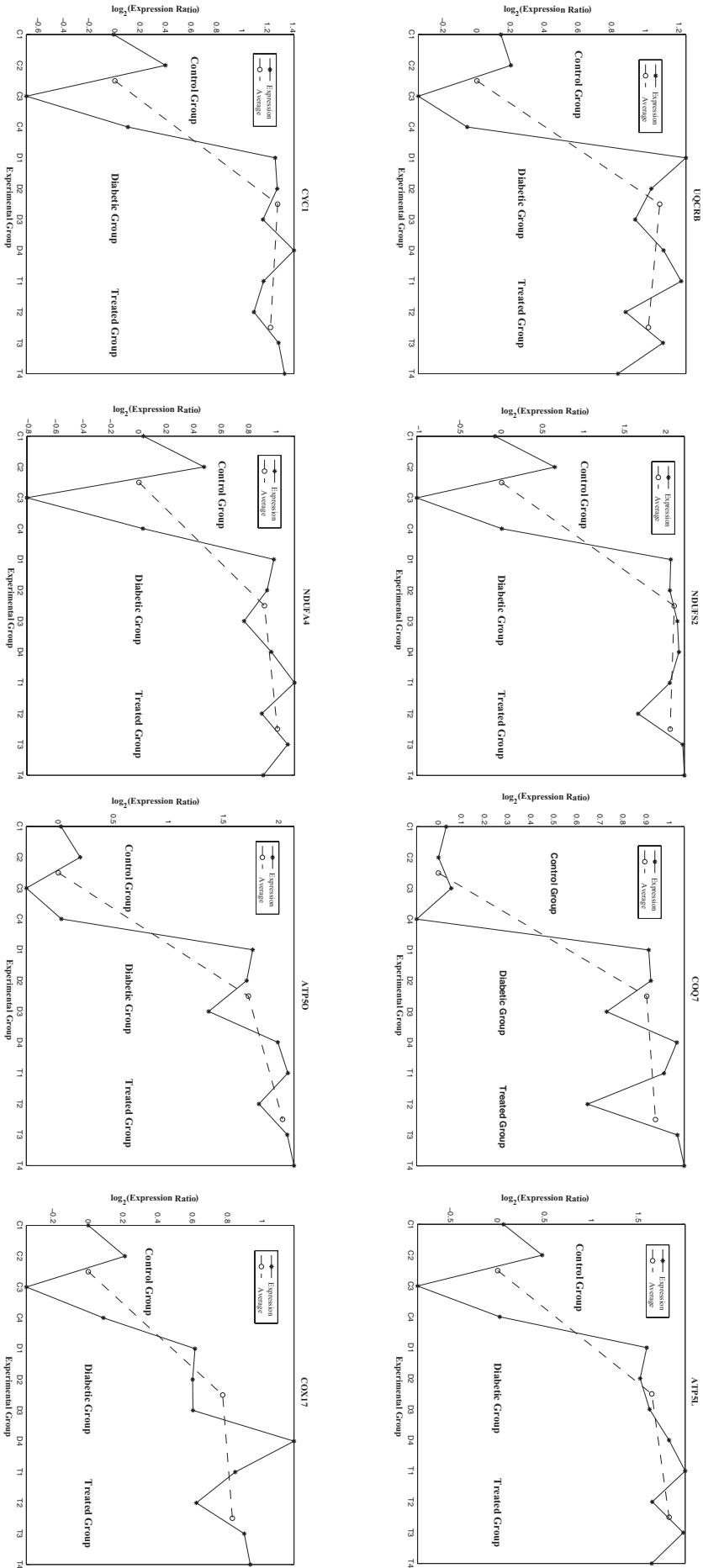


Figure 58  
Dysregulated genes annotated in oxidative phosphorylation pathway

Both datasets of genes, OXPHOS genes and fatty acid metabolism genes, contain a great amount of genes significant expressed in the diabetic and treated group. This denotes the dysregulation of these genes in the presence of the disease and the lack of a control after drug treatment.

The genes involved in fatty acid metabolism that present an expression dysregulation  $\geq 2$ -fold change do not observe a similar profile. Some genes are downregulated during the disease and others are upregulated. However, in neither case the gene returns to normal expression levels after treatment. In contrast, the genes involved in OXPHOS are upregulated in the diabetic mouse, with the exception of BCS1L. The expression of these genes is not affected by the presence of the drug since their expression in the treatment group does not change in comparison with the diabetic group.

The similar expression profiles of the OXPHOS genes suggest that the mitochondrial respiratory rate is increasing in the diabetic mouse. However, the lack of expression data for the mitochondrially encoded genes prevents from predicting a disruption of mitochondria activity.

## Chapter V : Materials and Methods

### 6.1 Experiment Design

Five to 6 week old male homozygous, obese KSJ *db/db* mice (C57BLKS- *mLepr db/db*) and its corresponding lean control the heterozygote *db/m* mice were obtained from Jackson Laboratories<sup>15</sup>. Four to five mice per cage were housed in micro-isolators. At week 17 the *db/db* mice were randomly assigned into two test groups: *db/db* untreated and *db/db* treated with PFD. The experiment design consisted in three different experimental groups: 1) the homozygote *db/db* mice or diabetic group, 2) the heterozygote *db/m* mice or control group, and 3) the *db/db* mice treated with pirfenidone (PFD) or treated group.

The *db/db* treated mice were individually caged and received oral PFD from week 17 to week 21, a total of 0.5% PFD was added to their diet, receiving an approximately intake of 25 mg/d of PFD. At the end of week 21 all the groups underwent 24 hours urine collection and later on were sacrificed. Left kidney of each mouse was snap-frozen in liquid nitrogen for RNA analysis. For a more detail information please refer to RamachandraRao, S. *et al.*, 2009.

The vivarium is located at the laboratories of the Center for Renal Translation Medicine, University of California, San Diego. All the mouse experiments practiced during the study were performed in accordance with the *NIH Guidelines for Research*

---

<sup>15</sup> <http://www.jax.org>

*Involving Recombinant DNA*<sup>16</sup> and the UCSD *Policy on the Use of Animals in Research and Teaching*<sup>17</sup> issued by the Institutional Animal Care and Use Committee, University of California, San Diego.

## 6.2 RNA Isolation and Analysis

Total RNA was extracted from mouse left kidneys using the RNeasy Mini Kit from Qiagen<sup>18</sup> following the manufacturer's protocol. Between 20-30 mg of starting material was used. cRNA was prepared and biotinylated using the Illumina TotalPrep RNA Amplification Kit from Ambion, Inc.<sup>19</sup> The cRNA was hybridized into the Illumina MouseRef-8 v2.0 Expression BeadChip from Illumina, Inc.<sup>20</sup> The microarray experiment was done at the Biomedical Genomics Laboratory (BIOGEM)<sup>21</sup>, University of California, San Diego.

## 6.3 Microarray Data Preliminary Filtration

The transcriptome analysis of the system was conducted across 4 different samples for each experimental group with a total of 45,281 different probes, resulting in 12 sets of microarray expression data for each probe. The data preparation consists of the following three steps: 1) only probes with available gene information were taken into consideration. 2) Average the expression values of those genes that have

---

<sup>16</sup> [http://oba.od.nih.gov/rdna/nih\\_guidelines\\_oba.html](http://oba.od.nih.gov/rdna/nih_guidelines_oba.html)

<sup>17</sup> <http://iacuc.ucsd.edu/regulations.aspx>

<sup>18</sup> <http://www.qiagen.com>

<sup>19</sup> <http://www.ambion.com>

<sup>20</sup> <http://www.illumina.com>

<sup>21</sup> <http://microarrays.ucsd.edu/>

more than one dataset. 3) Discard the genes that did not show an expression value above 100 in any of the 12 datasets. The final list used in the analysis of the expression profile was reduced to 5,589 probes with unique gene information.

## 6.4 Statistical Analysis and Gene Expression Clustering

Using the statistical tool Significance Analysis of Microarray (SAM)<sup>22</sup> (Tusher et al. 2001) the genes that have a statistically significant change in expression in the set of the microarray experiments were identified. SAM classified the genes as potentially significant genes based on a set of gene-specific *t* test and by assigning a score to each gene according on its change in expression relative to the standard deviation. Genes with a score greater than a threshold ( $\Delta$ ) are potentially significant genes. SAM normalized the data via simple media centering of the arrays.

For the present analysis it was established that the selected  $\Delta$  should not result in a false discovery rate (FDR) greater than 10%. The analysis was performed on a pairwise basis and the results of each individual analysis were compared to cluster the genes depending on the set of experiments where they showed a significant expression. The gene expression clusters are as followed: a) Control, b) Diabetic, c) Treated, d) Control and Diabetic, e) Control and Treated, f) Diabetic and Treated, and g) All Groups (Control, Diabetic, and Treated).

---

<sup>22</sup> <http://www-stat.stanford.edu/~tibs/SAM/>

## 6.5 Orthology

The dataset for PPI in mouse is reduced so a network-based analysis of the genes within a mouse PPI network will be insufficient. However, the information available for the human genome provides an adequate platform for this analysis and the study of any species can be made if human orthologs genes are available.

The human orthologs of the mouse genes used on the presente work were identified using the bioinformatics data management tool BioMart<sup>23</sup> (Smedley et al. 2009). BioMart provides homology by mapping specific genes across all the species present in Ensembl.

## 6.6 Protein-Protein Interaction Network

The human protein-protein interaction network was assembled with a dataset that consists of 11,262 genes with a total of 58,662 interactions. The dataset includes yeast two-hybrid experiments, predicted interaction via orthology, and literature curation and review. The human orthologs of the 7 different gene expression clusters were mapped onto the human PPI network using Cytoscape<sup>24</sup> version 2.6.3 (Cline et al. 2007).

---

<sup>23</sup> <http://www.biomart.org/>

<sup>24</sup> <http://www.cytoscape.org/>

## 6.7 Gene Ontology Enrichment Analysis

The enrichment analysis of the expression clusters was done employing the Database Annotation, Visualization and Integrated Discovery (DAVID)<sup>25</sup> v6.7 (Huang et al. 2009b; 2009a) a web-accessible server. DAVID's operation is based on a modified Fisher's Exact Test to calculate the statistical significance of the enrichment score of each group of genes that was analyzed. The background employed for the analysis was the list of genes that integrate the complete human PPI network.

## 6.8 Expression Analysis

For the analysis of the genes expression the expression ratio ( $ER$ ) of each gene in the array was calculated according to Equation 1.

$$ER_{ij} = \frac{EXP_{ij}}{\overline{C}_i} \quad \text{Equation 1}$$

The index  $i$  goes from 1 to  $N$  number of genes in the array and the index  $j$  goes from 1 to  $n$  number of experiments (in the case of the present work  $j = 12$ , since there are 4 experiments for each condition or experimental group).  $ER_{ij}$  is the expression ratio,  $EXP_{ij}$  is the array data, and  $\overline{C}_i$  is the average array data on the control group.

---

<sup>25</sup> <http://david.abcc.ncifcrf.gov/home.jsp>



In order to observe a symmetrical behavior of the expression the ratios were plotted using a *log2* scale. This way the up and downregulated genes are treated in a similar fashion.

The expression correlation between two genes was measured using the Pearson's Correlation Coefficient (*R*). The coefficient was calculated from the logarithm base 2 of the expression ratios of the genes employing the function *corrcoef()* from the numerical computing environment MatLab.

## 6.9 Fatty Acid Metabolism Pathway and Oxidative Phosphorylation

In order to obtain the genes those are annotated for fatty acid metabolism and oxidative phosphorylation two mitochondrial databases were employed. The Mitochondrial Protein Interactome Database – MitoInteractome<sup>26</sup> (Reja et al. 2009) and the Database for Mitochondria-Related Genes, Proteins and Diseases – Mitochondrial Proteome (MitoP2)<sup>27</sup> (Prokisch et al. 2006).

For the MitoP2 Database the criteria employed was: all the genes annotated for the functional category Respiratory Chain/Oxidative Phosphorylation and Lipid Metabolism. Only the genes with a SVM prediction score > 1 were considered significant, following the suggestions made by the developers.

---

<sup>26</sup> <http://mitointeractome.kobic.kr/>

<sup>27</sup> <http://www.mitop.de:8080/mitop2/>

For the MitoInteractome Database the search criteria were: all the genes annotated for the KEGG Pathways Oxidative Phosphorylation (KEGG ID: hsa00190) and Fatty Acid Metabolism (KEGG ID: hsa00071).

## Chapter VI : Final Remarks

### 7.1 Conclusion

The present work is a comprehensive analysis of the transcriptome in a mouse disease model of diabetes mellitus before and after treatment with the antifibrotic drug PFD. The analysis provides a wide perspective of the repercussions that the disease has over the transcription of the mouse genome and provides important clues about the mechanism of action of PFD. The main discussion moves around PGC-1 $\alpha$  due to three main reasons: 1) the expression of PGC-1 $\alpha$  is affected during stress conditions present in diabetic patients, such as, high content of free fatty acids and high concentrations of ROS. 2) PGC-1 $\alpha$  is a multifunctional molecule that regulates key biological pathways, such as, energy homeostasis and mitochondria activity that are disrupted in diabetic patients. 3) Recent studies associate single nucleotide polymorphism of PGC-1 $\alpha$  with insulin resistance and the pathogenesis of type 2 diabetes (Hara et al. 2002; Muller et al. 2003).

The statistical classification of the genes into expression clusters and the enrichment analysis of each group provide the first perspective of the expression behavior. The diabetic experimental group has significant expression of genes involved in stress response and p53 signaling pathway indicating that the transcriptome is responding to the stress conditions present during the development of the disease. The high concentration of ROS is a common characteristic under diabetic conditions. The activation of the p53 signaling pathway may lead to deterioration or

apoptosis of  $\beta$ -cells that are vulnerable to high concentration of ROS due to the low expression of antioxidant enzymes within this type of cells. Moreover, the enrichment of the diabetic and treated group reflects that the OXPHOS and the energy production are dysregulated as a consequence of the energy homeostasis is disrupted. These results are tightly connected since the respiratory chain is the main producer of ROS. The high content of free fatty acids, another characteristic of diabetes, possibly increases the mitochondrial respiratory rate that results in a high concentration of ROS and finally affecting the insulin production of  $\beta$ -cells. The ER and the proteasome complex also respond to stress conditions, in especial, to the oxidative stress. The fact that the treated and diabetic experimental group shows enrichment in protein metabolic process and negative regulation of gene expression suggest that the ER and the proteasome are responding to the disease. The ER activates the UPR reducing the transcription of genes and promoting the degradation of proteins by the ubiquitin-proteasome system. Although, these stress response are not specific to any type of protein the insulin signaling pathway might be the most impacted because the ubiquitination process affects the regulation of insulin signaling and its action. Furthermore the exposure of the cell to an increase content of insulin, hyperinsulinemia, promotes the degradation of IRS-1 by the ubiquitin-proteasome system.

The presence of the drug seems to have an impact in the processing of RNA and the production of mature RNA. The results show an enrichment of the control and treated experimental group in RNA processing and the genes that are part of this term clearly are dysregulated under the diabetic state and recover their expression levels after the implementation of the treatment. However, it is not clear how this is

related to other biological functions that show great alterations during the development of the disease. In addition, the MAPK signaling pathway and the protein kinase cascade are also enriched in the control and treated group, suggesting that the drug regulates the kinase cascade pathway after a possible disturbance under the diabetic condition. This is relevant to the study because by the activation of the MAPK pathway enhances antioxidant defenses and stimulates the expression of PGC-1 $\alpha$ .

The analysis of PGC-1 $\alpha$  and its first neighbors in the PPI network shows the multifunctional activity of this gene. The main function of PGC-1 $\alpha$  is to increase the rate of transcription by binding to transcription factors and facilitating the recruitment of the transcription machinery. However, the interactions of PGC-1 $\alpha$  are not limited to transcription factors and coactivators. It also has a significant amount of neighbors involved in the spliceosome event, such as, splicing, polyadenylation and cleavage factors. The data demonstrates that the genes involved in RNA splicing that are part of the PGC1- $\alpha$  PPI subnetwork, are dysregulated during the disease and have a tendency to return to normal expression levels after drug treatment. The expression of these genes is usually suppressed under the diabetic condition. The enrichment in RNA splicing suggests that PGC-1 $\alpha$  is not only a facilitator of transcription but also a regulator of the splicing event of the just transcribed RNA. This is supported by the correlation that exists between PGC-1 $\alpha$  with the splicing factors and other genes involved in the spliceosome event. The low expression of PGC-1 $\alpha$  could cause an uncontrolled splicing process that result in a low or incorrect protein production.

The analysis of PGC-1 $\alpha$  first neighbors involved in lipid metabolism and mitochondria biogenesis is less conclusive but provides important information. The expression of the peroxisome proliferator activated receptors (PPAR), in specific, PPAR $\alpha$ , does not change across the experimental groups. However, the low expression of RXR may contribute to a low activity of PPAR $\alpha$ . The expression of the ESRs and ESRRs, in particular, ESR1 and ESRR $\alpha$ , is downregulated in the diabetic state. The low expression of these two genes that share common biological functions can result in low expression of nuclear factors required for the transcription of nuclear encoded mitochondrial genes. However the expression of NRF-1 that is regulated by ESR1 does not have major changes across the experimental groups even though it has a tendency to be downregulated.

The NRF-1 and 2 regulate the expression of the transcription factors, TFAM, TF1BM and TF2BM, required for the transcription of mtDNA. TF1BM and TF2BM do not have a clear dysregulation. Nevertheless, TFAM is clearly upregulated in the diabetic state and the expression levels are not suppressed after treatment. The linear relation between the amount of TFAM and the amount of mtDNA suggests that the number of mitochondrion in the system is increasing. But the low expression of POLRMT may reduce transcriptional activity of the mitochondrial genome that can result in a diminish mitochondrial activity.

The genes involved in the fatty acid metabolism and OXPHOS that present significant dysregulation in most case are overexpressed in the diabetic state. Even though this suggests that the mitochondrial oxidative and respiratory functions are increasing, it cannot be taken as a conclusive result. The lack of expression profile

for the mitochondrially encoded genes limits the establishment of a concrete statement. However, the microarray data clearly shows that the drug does not affect the expression of genes involved in one of these two biological pathways.

In conclusion type 2 diabetes provokes the stimulation of stress response probably due to the presence of an oxidative intracellular environment caused by a high respiration and fatty acid oxidative rates. The stress environment activates stress-activated kinases, in special the MAPK pathway, that reduce insulin signaling and the cellular uptake of glucose. In the other hand, PGC-1 $\alpha$  besides increasing the transcription rates it might also play a regulatory role in the formation of mature mRNA. The expression inhibition of PGC-1 $\alpha$  under diabetic conditions may increase the splicing rates and the production of more proteins. The overproduction of proteins exceeds the capacity of the folding machinery as a reaction the unfolded protein response (UPR) is activated. The ongoing activity of the UPR induced ER-stress, increasing the concentration of ROS activating the kinases pathways. The regulation of RNA processing by PFD could be a result of the effect of the drug in the expression of PGC-1 $\alpha$ . The presence of the drug recovers the expression levels of PGC-1 $\alpha$  that was inhibited under diabetic conditions. The normal expression of PGC-1 $\alpha$  regulates the splicing mechanism and the cell returns to normal levels of protein production and reduces the protein folding demands. The ER-stress is inhibited and the generation of ROS is reduced.

## 7.2 Limitations and Future Work

The present analysis was based on expression data obtained from DNA microarray experiments. The interpretation of mRNA measurements assumes that there is a one-to-one relation between the mRNA concentration and the protein concentration. However the translation of DNA into mRNA and the transcription of mRNA into proteins are complex mechanisms. The translation and transcription processes are conformed by sequence of elementary steps and involves catalytic and reaction kinetics that grant great complexity to these central cellular processes (Mehra et al. 2003). Moreover the translation of individual mRNA into the corresponding protein is a regulated process. All these characteristics disrupt the one-to-one correlation between measured mRNA and protein concentration (MacKay et al. 2004) making it difficult to produce unquestionable results base solely in microarray data.

The principal limitation of the present work is the lack of experimental replicates. Even though a single replicate of microarray experiment provides enough information to generate initial hypothesis for further testing (Sásik et al. 2004) it is required to corroborate the results with successive microarray replicates or by quantifying gene expression with quantitative RT-PCR or Northern blotting.

The results presented on the current work should be experimentally verified in the future. The validation of the results could be done using a proteomic approach and by quantifying the expression of those genes that are fundamental in the regulation of key biological pathways. The current analysis provides the guidelines for the upcoming work and dictates the path to follow. The results furnish a



panoramic picture of what is happening within the cell and identify the crucial elements for the overall understanding of diabetes and diabetic nephropathy as complex events.

## REFERENCES

- Abrass, C.K. (1994) Diabetic Nephropathy. Mechanism of Mesangial Matrix Expansion. *The West Journal of Medicine*, 162 (4), pp. 318-21.
- Al-Nozha, M.M., Al-Maatouq, M.A., Al-Mazrou, Y.Y., Al-Harthi, S.S., Arafah, M.R., Khalil, M.Z., Khan, N.B., Al-Khadra, A., Al-Marzouki, K. & Nouh, M.S. (2004) Diabetes mellitus in Saudi Arabia. *Saudi medical journal*, 25 (11), pp.1603–1610.
- Alqurashi, K. A, Aljabri, K.S. & Bokhari, S. A. (2011) Prevalence of diabetes mellitus in a Saudi community. *Annals of Saudi medicine*, 31 (1), pp.19-23.
- American Diabetes Association (2011) Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, 34 (Supplement\_1), pp.S62-S69.
- American Diabetes Association (2008) Economic costs of diabetes in the U.S. In 2007. *Diabetes care*, 31 (3), pp.596-615.
- Anderson, J.W., Kendall, C.W.C. & Jenkins, D.J. (2003) Importance of weight management in type 2 diabetes: review with meta-analysis of clinical studies. *Journal of the American College of Nutrition*, 22 (5), pp.331-9.
- Andersson, U. & Scarpulla, R.C. (2001) Pgc-1-related coactivator, a novel, serum-inducible coactivator of nuclear respiratory factor 1-dependent transcription in mammalian cells. *Molecular and cellular biology*, 21 (11), pp.3738-49.
- Ansorge, W.J. (2009) Next-generation DNA sequencing techniques. *New biotechnology*, 25 (4), pp.195-203.
- Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A.T., Kerrien, S., Khadake, J., Kerssemakers, J., Leroy, C., Menden, M., Michaut, M., Montecchi-Palazzi, L., Neuhauser, S.N., Orchard, S., Perreau, V., Roechert, B., Eijk, K. van & Hermjakob, H. (2010) The

- IntAct molecular interaction database in 2010. *Nucleic acids research*, 38 (Database issue), pp.D525-31.
- Ariza, M.A., Vimalananda, V.G. & Rosenzweig, J.L. (2010) The economic consequences of diabetes and cardiovascular disease in the United States. *Reviews in endocrine & metabolic disorders*, 11 (1), pp.1-10.
- Asin-Cayuela, J. & Gustafsson, C.M. (2007) Mitochondrial transcription and its regulation in mammalian cells. *Trends in biochemical sciences*, 32 (3), pp.111-7.
- Attur, M.G., Dave, M.N., Tsunoyama, K., Akamatsu, M., Kobori, M., Miki, J., Abramson, S.B., Katoh, M. & Amin, a R. (2002) “A system biology” approach to bioinformatics and functional genomics in complex human diseases: arthritis. *Current issues in molecular biology*, 4 (4), pp.129-46.
- Balasubramanyam, M., Sampathkumar, R. & Mohan, V. (2005) Is insulin signaling molecules misguided in diabetes for ubiquitin-proteasome mediated degradation? *Molecular and cellular biochemistry*, 275 (1-2), pp.117-25.
- Barroso, I., Gurnell, M., Crowley, V.E., Agostini, M., Schwabe, J.W., Soos, M. a, Maslen, G.L., Williams, T.D., Lewis, H., Schafer, a J., Chatterjee, V.K. & O’Rahilly, S. (1999) Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*, 402 (6764), pp.880-3.
- Ben-Shachar, D. (2009) The interplay between mitochondrial complex I, dopamine and Sp1 in schizophrenia. *Journal of neural transmission (Vienna, Austria : 1996)*, 116 (11), pp.1383-96.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1), pp.289–300.
- Berggård, T., Linse, S. & James, P. (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics*, 7 (16), pp.2833-42.

- Bertalanffy, L.V. (1950) An Outline of General System Theory. *The British Journal for the Philosophy of Science*, I (2), pp.134-165.
- Breitkreutz, B.J., Stark, C., Reguly, T., Boucher, L., Breitkreutz, A., Livstone, M., Oughtred, R., Lackner, D.H., Bähler, J., Wood, V., Dolinski, K. & Tyers, M. (2008) The BioGRID Interaction Database: 2008 update. *Nucleic acids research*, 36 (Database issue), pp.D637-40.
- Brosius, F.C., Alpers, C.E., Bottinger, E.P., Breyer, M.D., Coffman, T.M., Gurley, S.B., Harris, R.C., Kakoki, M., Kretzler, M., Leiter, E.H., Levi, M., McIndoe, R. a, Sharma, K., Smithies, O., Susztak, K., Takahashi, N. & Takahashi, T. (2009) Mouse models of diabetic nephropathy. *Journal of the American Society of Nephrology : JASN*, 20 (12), pp.2503-12.
- Brown, K.R. & Jurisica, I. (2005) Online predicted human interaction database. *Bioinformatics (Oxford, England)*, 21 (9), pp.2076-82.
- Chan, E.Y. (2005) Advances in sequencing technology. *Mutation research*, 573 (1-2), pp.13-40.
- Chan, V., Graves, D.J. & McKenzie, S.E. (1995) The biophysics of DNA hybridization with immobilized oligonucleotide probes. *Biophysical journal*, 69 (6), pp.2243-55.
- Chang, Y.-C. & Chuang, L.-M. (2010) The role of oxidative stress in the pathogenesis of type 2 diabetes: from molecular mechanism to clinical implication. *American journal of translational research*, 2 (3), pp.316-31.
- Chatr-Aryamontri, A., Ceol, A., Palazzi, L.M., Nardelli, G., Schneider, M.V., Castagnoli, L. & Cesareni, G. (2007) MINT: the Molecular INTeraction database. *Nucleic acids research*, 35 (Database issue), pp.D572-4.
- Cheng, K., Yang, N. & Mahato, R.I. (2009) TGF-beta1 gene silencing for treating liver fibrosis. *Molecular pharmaceutics*, 6 (3), pp.772-9.

- Chisamore, M.J., Wilkinson, H.A., Flores, O. & Chen, J.D. (2009) Estrogen-related receptor-alpha antagonist inhibits both estrogen receptor-positive and estrogen receptor-negative breast tumor growth in mouse xenografts. *Molecular cancer therapeutics*, 8 (3), pp.672-81.
- Chuang, H.-Y., Hofree, M. & Ideker, T. (2010) A Decade of Systems Biology. *Annual review of cell and developmental biology*, 26, pp. 721-44.
- Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.-L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T. & Bader, G.D. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nature protocols*, 2 (10), pp.2366-82.
- Collas, P. (2010) The current state of chromatin immunoprecipitation. *Molecular biotechnology*, 45 (1), pp.87-100.
- Collins, F.S., Morgan, M. & Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, 300 (5617), pp.286-90.
- Djamali, A. & Samaniego, M. (2009) Fibrogenesis in Kidney Transplantation: Potential Targets for Prevention and Therapy. *Transplantation*, 88 (10), pp.1149-1156.
- Doolittle, J.M. & Gomez, S.M. (2010) Structural similarity-based predictions of protein interactions between HIV-1 and Homo sapiens. *Virology journal*, 7 (82).
- Doyle, M.A., Gasser, R.B., Woodcroft, B.J., Hall, R.S. & Ralph, S. a (2010) Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. *BMC genomics*, 11 (222).
- Essop, M., Chan, W. & Hattingh, S. (2010) Proteomic Analysis of Mitochondrial Proteins in a Mouse Model of Type 2 Diabetes. *Cardiovascular Journal of Africa*, 21, Online publication.

- Expert Committee on the Diagnosis and Classification of Diabetes Mellitus, The (2002) Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes care*, 25 (Suppl 1), pp.s5-20.
- Ferreira, E.N., Rangel, M.C.R., Galante, P.F., Souza, J.E. de, Molina, G.C., Souza, S.J. de & Carraro, D.M. (2010) Alternative splicing enriched cDNA libraries identify breast cancer-associated transcripts. *BMC genomics*, 11(Suppl 5), S4.
- Fields, S. & Song, O. (1989) A Novel Genetic System to Detect Protein-Protein Interactions. *Nature*, 340 (6230), pp.245-46.
- Finck, B.N. & Kelly, D.P. (2006) PGC-1 Coactivators: Inducible Regulators of Energy Metabolism in Health and Disease. *Journal of Clinical Investigation*, 116 (3), pp.615-622.
- Friedman, C.P., Altman, R.B., Kohane, I.S., McCormick, K.A., Miller, P.L., Ozbolt, J.G., Shortliffe, E.H., Stormo, G.D., Szczepaniak, M. & Tuck, D. (2004) Training the Next Generation of Informaticians: The Impact of “BISTI” and Bioinformatics—A Report from the American College of Medical Informatics. *Journal of the American Medical Informatics Association*, 11 (3), pp.167-72.
- Froguel, P., Vaxillaire, M., Sun, F., Velho, G., Zouali, H., Butel, M., Lesage, S., Vionnet, N., Clement, K. & Fougereuse, F. (1992) Close Linkage of Glucokinase Locus on Chromosome 7p to Early-Onset Non-Insulin-Dependent Diabetes Mellitus. *Nature*, 356 (6365), pp.162–164.
- Fruchart, J.-C. (2009) Peroxisome proliferator-activated receptor-alpha (PPARalpha): at the crossroads of obesity, diabetes and cardiovascular disease. *Atherosclerosis*, 205 (1), pp.1-8.
- Fukuoh, A., Ohgaki, K., Hatae, H., Kuraoka, I., Aoki, Y., Uchiumi, T., Jacobs, H.T. & Kang, D. (2009) DNA conformation-dependent activities of human mitochondrial RNA polymerase. *Genes to cells : devoted to molecular & cellular mechanisms*, 14 (8), pp.1029-42.
- Gan, Y., Herzog, E.L. & Gomer, R.H. (2011) Pirfenidone treatment of idiopathic pulmonary fibrosis. *Therapeutics and clinical risk management*, 7, pp.39-47.

- Garstka, H.L. (2003) Import of mitochondrial transcription factor A (TFAM) into rat liver mitochondria stimulates transcription of mitochondrial DNA. *Nucleic Acids Research*, 31 (17), pp.5039-5047.
- Gerstein, M., Greenbaum, D., Cheung, K. & Miller, P.L. (2007) An interdepartmental Ph.D. program in computational biology and bioinformatics: the Yale perspective. *Journal of biomedical informatics*, 40 (1), p.pp.73-9.
- Gleyzer, N., Vercauteren, K. & Scarpulla, R.C. (2005) Control of mitochondrial transcription specificity factors (TFB1M and TFB2M) by nuclear respiratory factors (NRF-1 and NRF-2) and PGC-1 family coactivators. *Molecular and cellular biology*, 25 (4), pp.1354-66.
- Grieco, F.A., Vendrame, F., Spagnuolo, I. & Dotta, F. (2011) Innate immunity and the pathogenesis of type 1 diabetes. *Seminars in immunopathology*, 33 (1), pp.57-66.
- Ha, H., Hwang, I.-A., Park, J.H. & Lee, H.B. (2008) Role of reactive oxygen species in the pathogenesis of diabetic nephropathy. *Diabetes research and clinical practice*, 82 Suppl 1, pp.S42-5.
- Handschin, C. & Spiegelman, B. (2006) Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism. *Endocrine reviews*, 27 (7), pp.728-35.
- Haneda, M., Koya, D., Isono, M. & Kikkawa, R. (2003) Overview of glucose signaling in mesangial cells in diabetic nephropathy. *Journal of the American Society of Nephrology*, 14, pp. 1374-82.
- Hara, K., Tobe, K., Okada, T., Kadowaki, H., Akanuma, Y., Ito, C., Kimura, S. & Kadowaki, T. (2002) A genetic variation in the PGC-1 gene could confer insulin resistance and susceptibility to Type II diabetes. *Diabetologia*, 45 (5), pp.740-3.

- Hocquette, J.F. (2005) Where are we in genomics? *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society*, 56 Suppl 3, pp.37-70.
- Hsu, Y.-C., Chen, M.-J., Yu, Y.-M., Ko, S.-Y. & Chang, C.-C. (2010) Suppression of TGF- $\beta$ 1/SMAD pathway and extracellular matrix production in primary keloid fibroblasts by curcuminoids: its potential therapeutic use in the chemoprevention of keloid. *Archives of dermatological research*, 302 (10), pp.717-24.
- Hsueh, W., Abel, E.D., Breslow, J.L., Maeda, N., Davis, R.C., Fisher, E.A., Dansky, H., McClain, D. a, McIndoe, R., Wassef, M.K., Rabadán-Diehl, C. & Goldberg, I.J. (2007) Recipes for creating animal models of diabetic cardiovascular disease. *Circulation research*, 100 (10), pp.1415-27.
- Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009a) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, 37 (1), pp.1-13.
- Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009b) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4 (1), pp.44-57.
- Hubbard, T.J.P., Aken, B.L., Beal, K., Ballester, B & Caccamo, M. (2007) Ensembl 2007. *Nucleic acids research*, 35 (Database issue), pp.D610-7.
- Ideker, T., Galitski, T. & Hood, L. (2001) A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2, pp.343-72.
- International Diabetes Federation (2009) IDF Diabetes Atlas. *International Diabetes Federation*, 4<sup>th</sup> Edition.
- Jazet, I.M., Pijl, H. & Meinders, A.E. (2003) Adipose tissue as an endocrine organ: impact on insulin resistance. *The Netherlands journal of medicine*, 61 (6), pp.194-212.



- Jovanovic, L. (2009) Definition, size of the problem, screening and diagnostic criteria: who should be screened, cost-effectiveness, and feasibility of screening. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 104 Suppl , pp.S17-9.
- Kadowaki, T., Bevins, C.L., Cama, A., Ojamaa, K., Marcus-Samuels, B., Kadowaki, H., Beitz, L., McKeon, C. & Taylor, S.I. (1988) Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. *Science*, 240 (4853), pp.787-90.
- Karantzali, E., Lekakis, V., Ioannou, M., Hadjimichael, C., Papamatheakis, J. & Kretsovali, A. (2010) Sall1 regulates embryonic stem cell differentiation in association with Nanog. *The Journal of biological chemistry*, 286 (2), pp.1037-1045.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R. & Pandey, A. (2009) Human Protein Reference Database--2009 update. *Nucleic acids research*, 37 (Database issue), pp.D767-72.
- Kitano, H. (2002a) Computational systems biology. *Nature*, 420 (6912), pp.206-10.
- Kitano, H. (2002b) Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Current genetics*, 41 (1), pp.1-10.
- Klinge, C.M. (2008) Estrogenic control of mitochondrial function and biogenesis. *Journal of cellular biochemistry*, 105 (6), pp.1342-51.
- Koppen, A. & Kalkhoven, E. (2010) Brown vs white adipocytes: The PPARgamma coregulator story. *FEBS letters*, 584, pp.3250-3259.

- Krug, K., Nahnsen, S. & Macek, B. (2010) Mass spectrometry at the interface of proteomics and genomics. *Molecular bioSystems*, pp.284-291.
- Kubo, M., Ijichi, N., Ikeda, K., Horie-Inoue, K., Takeda, S. & Inoue, S. (2009) Modulation of adipogenesis-related gene expression by estrogen-related receptor gamma during adipocytic differentiation. *Biochimica et biophysica acta*, 1789 (2), pp.71-7.
- Lammers, G., Gilissen, C., Nillesen, S.T.M., Uijtdewilligen, P.J.E., Wismans, R.G., Veltman, J. a, Daamen, W.F. & Kuppevelt, T.H. van (2010) High density gene expression microarrays and gene ontology analysis for identifying processes in implanted tissue engineering constructs. *Biomaterials*, 31 (32), pp.8299-312.
- Li, M., Fu, W. & Li, X.-A. (2010) Differential fatty acid profile in adipose and non-adipose tissues in obese mice. *International journal of clinical and experimental medicine*, 3 (4), pp.303-7.
- Liang, H. & Ward, W.F. (2006) PGC-1alpha: a key regulator of energy metabolism. *Advances in physiology education*, 30 (4), pp.145-51.
- Lin, J., Puigserver, P., Donovan, J., Tarr, P. & Spiegelman, B. (2002) Peroxisome proliferator-activated receptor gamma coactivator 1beta (PGC-1beta), a novel PGC-1-related transcription coactivator associated with host cell factor. *The Journal of biological chemistry*, 277 (3), pp.1645-8.
- Liu, D., Zhang, Zhiping, Gladwell, W. & Teng, C.T. (2003) Estrogen stimulates estrogen-related receptor alpha gene expression through conserved hormone response elements. *Endocrinology*, 144 (11), pp.4894-904.
- Liu, Z.-P., Wang, Y., Zhang, X.-S., Xia, W. & Chen, L. (2011) Detecting and analyzing differentially activated pathways in brain regions of Alzheimer's disease patients. *Molecular BioSystems*.

- Livshits, M. A. & Mirzabekov, A. D. (1996) Theoretical analysis of the kinetics of DNA hybridization with gel-immobilized oligonucleotides. *Biophysical journal*, 71 (5), pp.2795-801.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M. & Horton, H. (1996) Expression Monitoring by Hybridization to High-density Oligonucleotide Arrays. *Nature biotechnology*, 14 (13), pp.1675–1680.
- Lowell, B.B. & Shulman, G.I. (2005) Mitochondrial dysfunction and type 2 diabetes. *Science*, 307 (5708), pp.384-7.
- MacKay, V.L., Li, X., Flory, M.R., Turcott, E., Law, G.L., Serikawa, K. A., Xu, X.L., Lee, H., Goodlett, D.R., Aebersold, R., Zhao, L.P. & Morris, D.R. (2004) Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Molecular & cellular proteomics : MCP*, 3 (5), pp.478-89.
- Macías-Barragán, J., Sandoval-Rodríguez, A., Navarro-Partida, J. & Armendáriz-Borunda, J. (2010) The multifaceted role of pirfenidone and its novel targets. *Fibrogenesis & tissue repair*, 3 (16).
- Maldonado, A.Y., Burz, D.S. & Shekhtman, A. (2010) In-cell NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*.
- Matsushima, Y., Adán, C., Garesse, R. & Kaguni, L.S. (2005) Drosophila mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells. *The Journal of biological chemistry*, 280 (17), pp.16815-20.
- McCarthy, M.I. & Hattersley, A.T. (2008) Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. *Diabetes*, 57 (11), pp.2889-98.

- Mehra, A., Lee, K.H. & Hatzimanikatis, V. (2003) Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. *Biotechnology and bioengineering*, 84 (7), pp.822-33.
- Meigs, J.B. (2009) Prediction of type 2 diabetes: the dawn of polygenetic testing for complex disease. *Diabetologia*, 52 (4), pp.568-70.
- Mensink, M., Blaak, E.E., Baak, M. a van, Wagenmakers, a J. & Saris, W.H. (2001) Plasma free Fatty Acid uptake and oxidation are already diminished in subjects at high risk for developing type 2 diabetes. *Diabetes*, 50 (11), pp.2548-54.
- Michnick, S.W., Ear, P.H., Manderson, E.N., Remy, I. & Stefan, E. (2007) Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nature reviews. Drug discovery*, 6 (7), pp.569-82.
- Mika, S. & Rost, B. (2006) Protein-protein interactions more conserved within species than across species. *PLoS computational biology*, 2 (7).
- Miller, J.A., Horvath, S. & Geschwind, D.H. (2010) Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (28), pp.12698-703.
- Mirebeau-Prunier, D., Le Pennec, S., Jacques, C., Gueguen, N., Poirier, J., Malthiery, Y. & Savagner, F. (2010) Estrogen-related receptor alpha and PGC-1-related coactivator constitute a novel complex mediating the biogenesis of functional mitochondria. *The FEBS journal*, 277 (3), pp.713-25.
- Mirsaidov, U.M., Wang, D., Timp, W. & Timp, G. (2010) Molecular diagnostics for personal medicine using a nanopore. *Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology*, 2 (4), pp.367-81.
- Moresco, J.J., Carvalho, P.C. & Yates, J.R. (2010) Identifying components of protein complexes in *C. elegans* using co-immunoprecipitation and mass spectrometry. *Journal of proteomics*, 73 (11), pp.2198-204.

- Morozova, O. & Marra, M. a (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92 (5), pp.255-64.
- Mullen, E. & Ohlendieck, K. (2010) Proteomic profiling of non-obese type 2 diabetic skeletal muscle. *International journal of molecular medicine*, 25 (3), pp.445–58.
- Muller, Y.L., Bogardus, C., Pedersen, O. & Baier, L. (2003) A Gly482Ser Missense Mutation in the Peroxisome Proliferator–Activated Receptor  $\gamma$  Coactivator-1 Is Associated With Altered Lipid Oxidation and Early Insulin Secretion in Pima Indians. *Diabetes*, 52 (3), pp.895-98.
- National Diabetes Information Clearinghouse (2007). Monogenic Forms of Diabetes: Neonatal Diabetes Mellitus and Maturity-onset Diabetes of the Young. *National Institute of Health*, Publication No. 07-6141.
- Ostlund, G., Schmitt, T., Forslund, K., Köstler, T., Messina, D.N., Roopra, S., Frings, O. & Sonnhammer, E.L.L. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic acids research*, 38 (Database issue), pp.D196-203.
- Oyadomari, S., Araki, E. & Mori, M. (2002) Endoplasmic reticulum stress-mediated apoptosis in pancreatic beta-cells. *Apoptosis : an international journal on programmed cell death*, 7 (4), pp.335-45.
- Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Mark, P., Stümpflen, V., Mewes, H.-W., Ruepp, A. & Frishman, D. (2005) The MIPS mammalian protein-protein interaction database. *Bioinformatics (Oxford, England)*, 21 (6), pp.832-4.
- Poncelet, a C., Caestecker, M.P. de & Schnaper, H.W. (1999) The transforming growth factor-beta/SMAD signaling pathway is present and functional in human mesangial cells. *Kidney international*, 56 (4), pp.1354-65.
- Prokisch, H., Andreoli, C., Ahting, U., Heiss, K., Ruepp, a, Scharfe, C. & Meitinger, T. (2006) MitoP2: the mitochondrial proteome database--now including mouse data. *Nucleic acids research*, 34 (Database issue), pp.D705-11.

- Puigserver, P. (2003) Peroxisome Proliferator-Activated Receptor-gamma Coactivator 1alpha (PGC-1alpha): Transcriptional Coactivator and Metabolic Regulator. *Endocrine Reviews*, 24 (1), pp.78-90.
- Puigserver, P., Wu, Z., Park, C.W., Graves, R., Wright, M. & Spiegelman, B. (1998) A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell*, 92 (6), pp.829-39.
- Raghu, G., Johnson, W.C., Lockhart, D. & Mageto, Y. (1999) Treatment of idiopathic pulmonary fibrosis with a new antifibrotic agent, pirfenidone: results of a prospective, open-label Phase II study. *American journal of respiratory and critical care medicine*, 159, pp.1061-9.
- RamachandraRao, S.P., Talwar, P., Ravasi, T. & Sharma, K. (2010) Novel systems biology insights using antifibrotic approaches for diabetic kidney disease. *Expert Review of Endocrinology and Metabolism*, 5 (1), pp.127–35.
- RamachandraRao, S.P., Zhu, Y., Ravasi, T., McGowan, T. a, Toh, I., Dunn, S.R., Okada, S., Shaw, M. a & Sharma, K. (2009) Pirfenidone is renoprotective in diabetic kidney disease. *Journal of the American Society of Nephrology : JASN*, 20 (8), pp.1765-75.
- Rasouli, N. & Kern, P. A. (2008) Adipocytokines and the metabolic complications of obesity. *The Journal of clinical endocrinology and metabolism*, 93 (11 Suppl 1), pp.S64-73.
- Razick, S., Magklaras, G. & Donaldson, I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC bioinformatics*, 9 (405).
- Rees, D.A. & Alcolado, J.C. (2005) Animal models of diabetes mellitus. *Diabetic medicine : a journal of the British Diabetic Association*, 22 (4), pp.359-70.
- Reimers, M. (2005) Statistical analysis of microarray data. *Addiction biology*, 10 (1), pp.23-35.

- Reja, R., Venkatakrishnan, A.J., Lee, J., Kim, B.-C., Ryu, J.-W., Gong, S., Bhak, J. & Park, D. (2009) MitoInteractome: mitochondrial protein interactome database, and its application in “aging network” analysis. *BMC genomics*, 10 (Suppl 3), S20.
- RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium, The (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, 409, pp.685-90.
- Rollerova, E. & Urbancikova, M. (2000) Intracellular estrogen receptors, their characterization and function (Review). *Endocrine regulations*, 34 (4), pp.203-18.
- Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P. & Vidal, M. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437 (7062), pp.1173-8.
- Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. & Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic acids research*, 32 (Database issue), pp.D449-51.
- Sanchez, A.P. & Sharma, K. (2009) Transcription factors in the pathogenesis of diabetic nephropathy. *Expert reviews in molecular medicine*, 11 (e13).
- Scarpulla, R.C. (2008a) Nuclear control of respiratory chain expression by nuclear respiratory factors and PGC-1-related coactivator. *Annals of the New York Academy of Sciences*, 1147, pp.321-34.
- Scarpulla, R.C. (1996) Nuclear Respiratory Factors and the Pathways of Nuclear-Mitochondrial Interaction. *Trends in Cardiovascular Medicine*, 6 (2), pp.39-45.
- Scarpulla, R.C. (2008b) Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiological reviews*, 88 (2), pp.611-38.

- Schofield, P.N., Eppig, J., Huala, E., De Angelis, M.H., Harvey, M., Davidson, D., Weaver, T., Brown, S., Smedley, D. & Rosenthal, N. (2010) Sustaining the data and bioresource commons. *Science*, 330 (6004), pp.592-93.
- Schulman, I.G., Shao, G. & Heyman, R.A. (1998) Transactivation by Retinoid X Receptor–Peroxisome Proliferator-Activated Receptor  $\gamma$  (PPAR $\gamma$ ) Heterodimers: Intermolecular Synergy Requires Only the PPAR $\gamma$  Hormone-Dependent Activation Function. *Molecular and cellular biology*, 18 (6), pp.3483-94.
- Sennblad, B. & Lagergren, J. (2009) Probabilistic orthology analysis. *Systematic biology*, 58 (4), pp.411-24.
- Shameli, A., Yamanouchi, J., Thiessen, S. & Santamaria, P. (2007) Endoplasmic reticulum stress caused by overexpression of islet-specific glucose-6-phosphatase catalytic subunit-related protein in pancreatic Beta-cells. *The review of diabetic studies : RDS*, 4 (1), pp.25-32.
- Sharma, K., McCue, P. & Dunn, S.R. (2003) Diabetic kidney disease in the *db/db* mouse. *American Journal of Physiology, Renal Physiology*, 284 (6), pp. F1138-44.
- Shaw, J.E., Sicree, R. a & Zimmet, P.Z. (2010) Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes research and clinical practice*, 87 (1), pp.4-14.
- Shen, Y.Y., Peake, P.W. & Charlesworth, J. a (2008) Review article: Adiponectin: its role in kidney disease. *Nephrology*, 13 (6), pp.528-34.
- Shihab, F.S. (2007) Do we have a pill for renal fibrosis? *Clinical journal of the American Society of Nephrology : CJASN*, 2 (5), pp.876-8.
- Shumway, J.T. & Gambert, S.R. (2002) Diabetic nephropathy-pathophysiology and management. *International urology and nephrology*, 34 (2), pp.257-64.



- Smedley, D., Haider, S., Ballester, Benoit, Holland, R., London, D., Thorisson, G. & Kasprzyk, A. (2009) BioMart--biological queries made easy. *BMC genomics*, 10 (22).
- Snijder, M. B., Heine, R. J., Seidell, J. C., Bouter, L. M., Stehouwer, C. D. A., Nijpels, G., et al. (2006). Associations of adiponectin levels with incident impaired glucose metabolism and type 2 diabetes in older men and women: the hoorn study. *Diabetes care*, 29(11), pp.2498-503.
- Srinivasan, K. & Ramarao, P. (2007) Animal models in type 2 diabetes research: an overview. *The Indian journal of medical research*, 125 (3), pp.451-72.
- Stockand, J.D. & Sansom, S.C. (1997) Regulation of filtration rate by glomerular mesangial cells in health and diabetic renal disease. *American Journal of Kidney Disease*, 29 (6), pp. 971-81.
- Stockand, J.D. & Sansom, S.C. (1998) Glomerular mesangial cells: electrophysiology and regulation of contraction. *Physiological Reviews*, 78 (3), pp. 723-44.
- Stumvoll, M., Goldstein, B.J. & Haeften, T.W. van (2005) Type 2 diabetes: principles of pathogenesis and therapy. *Lancet*, 365 (9467), pp.1333-46.
- Sásik, R., Woelk, C.H. & Corbeil, J. (2004) Microarray truths and consequences. *Journal of molecular endocrinology*, 33 (1), pp.1-9.
- Tamrakar, A.K., Singh, A.B. & Srivastava, A.K. (2009) db/+ Mice as an alternate model in antidiabetic drug discovery research. *Archives of medical research*, 40 (2), p.pp.73-8.
- Thibodeaux, G.N., Cowmeadow, R., Umeda, A. & Zhang, Z. (2009) A Tetracycline Repressor-based Mammalian Two-hybrid System to Detect Protein-protein Interactions in vivo. *Analytical biochemistry*, 386 (1), pp.129–131.

- Tulpan, D., Andronescu, M. & Leger, S. (2010) Free energy estimation of short DNA duplex hybridizations. *BMC bioinformatics*, 11 (105).
- Tusher, V.G., Tibshirani, R. & Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (9), pp.5116-21.
- Wang, Y.C. & Chen, B.S. (2011) A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC medical genomics*, 4 (2).
- Wanrooij, S. & Falkenberg, M. (2010) The human mitochondrial replication fork in health and disease. *Biochimica et biophysica acta*, 1797 (8), pp.1378-88.
- Wild, S., Roglic, G., Green, A., Sicree, R. & King, H. (2004) Global Prevalence of Diabetes. Estimates for the year 2000 and projections for 2030. *Diabetes care*, 27, pp.1047-53.
- Yamagata, K., Oda, N., Kaisaki, P.J., Menzel, S., Furuta, H., Vaxillaire, M., Southam, L., Cox, R.D., Lathrop, G.M. & Boriraj, V.V. (1996) Mutations in the Hepatocyte Nuclear Factor-1 $\alpha$  Gene in Maturity-Onset Diabetes of the Young (MODY3). *Nature*, 384 (6608), pp.455–458.
- Yang, Y., Mo, X., Chen, S., Lu, X. & Gu, D. (2011) Association of peroxisome proliferator-activated receptor gamma coactivator 1 alpha (PPARGC1A) gene polymorphisms and type 2 diabetes mellitus: a meta-analysis. *Diabetes/Metabolism Research and Reviews*, 27 (2), pp.177–184.
- Yellaboina, S., Dudekula, D.B. & Ko, M.S. (2008) Prediction of evolutionarily conserved interologs in *Mus musculus*. *BMC genomics*, 9 (465).
- Yessoufou, a & Wahli, W. (2010) Multifaceted roles of peroxisome proliferator-activated receptors (PPARs) at the cellular and whole organism levels. *Swiss medical weekly*, 140 (w13071).
- Yoshida, S., Tanaka, H., Oshima, H., Yamazaki, T., Yonetoku, Y., Ohishi, T., Matsui, T. & Shibasaki, M. (2010) AS1907417, a novel GPR119 agonist, as an

insulinotropic and  $\beta$ -cell preservative agent for the treatment of type 2 diabetes. *Biochemical and biophysical research communications*, 400 (4), pp.745-51.

Zhang, G., Lukoszek, R., Mueller-Roeber, B. & Ignatova, Z. (2010) Different sequence signatures in the upstream regions of plant and animal tRNA genes shape distinct modes of regulation. *Nucleic acids research*, pp.1-9.